



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 99026**

**TO: Ashwin Mehta  
Location: CM1/9E07/9E12  
Art Unit: 1638  
Monday, July 28, 2003**

**Case Serial Number: 09/991458**

**From: Toby Port  
Location: Biotech-Chem Library  
CM1-6A04  
Phone: 308-3534**

**toby.port@uspto.gov**

### **Search Notes**

Dear Examiner Mehta,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

**THIS PAGE BLANK (USPTO)**



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 – Circ. Desk



**THIS PAGE BLANK (USPTO)**



STIC-Biotech/ChemLib

99026

Fr m: Mehta, Ashwin  
Sent: Wednesday, July 16, 2003 5:50 PM  
To: STIC-Biotech/ChemLib  
Subject: seq search

RECEIVED

JUL 17 2003

STIC-BIOTECH/CHM LIB  
(STIC)

STIC,

Please search the commercial and interference databases for the following from 09/991,458:

- 1) the nucleotide sequence of SEQ ID NO:1
- 2) the amino acid sequence of SEQ ID NO: 2

My mail room is 9E12, office 9E07, art unit 1638.

Thank you,  
Ashwin

Ashwin Mehta  
United States Patent and Trademark Office  
Biotechnology Patent Examiner  
703-306-4540

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 22:33:23 ; Search time 7263 Seconds

(without alignments)  
11460.006 Million cell updates/sec

Title: US-09-991-458-1

Perfect score: 2860  
Sequence: 1 ggcacgaggtactgtgttg.....ataaaaaaaaaaaaaaaaaa 2860Scoring table:  
IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

```
GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vtl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2860	100.0	2860	3	AF013963	AF013963 Manduca s
2	2423.4	84.7	2861	3	AF006063	AF006063 Manduca s
3	446.6	15.6	4113	3	AY060329	AY060329 Drosophila
4	391.8	13.7	2473	6	AX191697	AX191697 Sequence
5	391.8	13.7	2473	6	AX121662	AX121662 Drosophila
6	383.4	13.4	2030	6	AX191701	AX191701 Sequence
7	382.8	13.4	2030	6	AX071318	AX071318 Drosophila
8	378.6	13.2	2097	3	AY058396	AY058396 Drosophila
9	378.6	13.2	2108	6	AX191703	AX191703 Sequence
10	366.8	12.8	1996	3	AF369383	AF369383 Aedes aeg
11	274.6	9.6	66809	2	AC015177	AC015177 Drosophila
12	274.6	9.6	166854	3	AC007330	AC007330 Drosophila
13	274.6	9.6	196337	3	AC005894	AC005894 Drosophila
14	274.6	9.6	242172	3	AE003832	AE003832 Drosophila
15	270.2	9.4	18567	2	AC018319	AC018319 Drosophila
16	270.2	9.4	16929	3	AC091203	AC091203 Drosophila
17	270.2	9.4	176929	3	AC091203	AC091203 Drosophila
18	270.2	9.4	286784	3	AE003352	AE003352 Drosophila
19	245.8	8.6	17325	2	AC018164	AC018164 Drosophila
20	245.8	8.6	184621	3	AC007440	AC007440 Drosophila
21	245.8	8.6	233148	3	AE003822	AE003822 Drosophila
22	231.4	8.1	37319	2	AC014400	AC014400 Drosophila
23	231.4	8.1	167902	3	AC014604	AC014604 Drosophila
24	231.4	8.1	302786	3	AE003434	AE003434 Drosophila
25	229.2	8.0	70610	2	AC0917561	AC0917561 Drosophila
26	229.2	8.0	153352	3	AC093103	AC093103 Drosophila
27	229.2	8.0	202495	3	AC092246	AC092246 Drosophila
28	229.2	8.0	260967	3	AE003645	AE003645 Drosophila
29	229.2	8.0	303043	3	DROSADH05	DROSADH05 Drosophila
30	227.2	7.9	1890	5	AF318177	AF318177 Drosophila
31	225.2	7.9	3375	10	AF411042	AF411042 Mus muscu
32	222	7.8	2817	6	AR048966	AR048966 Sequence
33	222	7.8	2817	6	AR080685	AR080685 Sequence
34	222	7.8	2817	10	RATGLY	L21672 Rat mRNA se
35	217.4	7.6	1894	9	AF286026	AF286026 Macaca mu
36	216	7.6	2255	4	BT099198	BT099198 Bos taurus
37	214.6	7.5	3393	10	AF109072	AF109072 Mus muscu
38	214.4	7.5	3188	4	BT099198	BT099198 Bos taurus
39	213.4	7.5	3738	9	AK096607	AK096607 Homo sapi
40	213	7.4	1873	10	AF109391	AF109391 Mus muscu
41	213	7.4	1888	10	MMU238309	MMU238309 Mus muscu
42	212.6	7.4	1985	10	RATDPRSP	M80233 Rat cocaine
43	211.8	7.4	1911	9	S80071	S80071 hPROR-brain
44	211	7.4	3404	10	RATDOPER	M80570 Rat dopamine
45	209.6	7.3	2722	10	RATLPTHA	M88111 Rattus norv

## ALIGNMENTS

```
RESULT 1
AF013963      2860 bp  mRNA  linear  INV 30-OCT-2001
LOCUS        Manduca sexta amino acid transporter/amino acid-gated channel for
DEFINITION  sodium/potassium ions (CAATCH1) mRNA, complete cds.
ACCESSION   AF013963
VERSION     AF013963.2  GI:16356923
KEYWORDS
SOURCE      Manduca sexta.
ORGANISM    Manduca sexta.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Sphingioidea; Sphingidae; Sphinginae; Manduca.
REFERENCE   1 (bases 1 to 2860)
AUTHORS     Feldman,D.H., Harvey,W.R. and Stevens,B.R.
```

```

gene      1. .2860
          /gene="CAATCH1"
CDS       101. .2002

```

Query Match	100.0%	Score 2860;	DB 3;	Length 2860;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2860; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

OY 1 GGACAGAGGTTACTTGTGTGGAGCAACTGTTTGGCGGTGTGCATCGATTTCTGAAACACA 60  
Db 1 GGACAGAGGTTACTTGTGTGGAGCAACTGTTTGGCGGTGTGCATCGATTTCTGAAACACA 60  
OY 61 ATACACCTTAACACATTCGCAAGTGTGATATTGTGGACCAAAATGATATGACGGCCCAAGTAA 120  
Db 61 ATACACCTTAACACATTCGCAAGTGTGATATTGTGGACCAAAATGATATGACGGCCCAAGTAA 120  
OY 121 CGGCGGTTTCGAGTGTGTGGAGCCCAAGATGGAACCAAAAGCATCATCGCAAAATTAAGTTT 180  
Db 121 CGGCGGTTTCGAGTGTGTGGAGCCCAAGATGGAACCAAAAGCATCATCGCAAAATTAAGTTT 180  
OY 181 ACCTCAGCAAAATTAATAAGCGGCGTCTCGATTAATATGATGACACGCACTTAAGGCTGA 240

Db	181	ACCCGACGAAATTAATAAGGCGCTGATTAATATATGATGACAGGCACTTAGAGCTGA	240
OY	241	ACGGCGAAGACGTATGGTATGTCACAAACATTGAATTGTTGATGTCGCATCGCTAC	300
Db	241	ACGGCGAAGACGTATGGTATGTCACAAACATTGAATTGTTGATGTCGCATCGCTAC	300
OY	301	ATCCGTCGTTTTGGGTAAACGTGTGGCGGTTCCCTTTCATGCGCAGACAAATGAGGAGG	360
Db	301	ATCCGTCGTTTTGGGTAAACGTGTGGCGGTTCCCTTTCATGCGCAGACAAATGAGGAGG	360
OY	361	TGCTTTCCTGTCGCATACGTACATCGTCTTTTACTTGTGCGCAAGCCTGTGTACTTT	420
Db	361	TGCTTTCCTGTCGCATACGTACATCGTCTTTTACTTGTGCGCAAGCCTGTGTACTTT	420
OY	421	AGAGTGTGCTCGGCAATTCAGTTCAGAAACCTCGTTAAAGTTTGGTCAATTTACC	480
Db	421	AGAGTGTGCTCGGCAATTCAGTTCAGAAACCTCGTTAAAGTTTGGTCAATTTACC	480
OY	481	GGCCATGAAAGGTACTGATACGCTCAAGTCGCGGCTCGGTTACATCCTGCTTACTA	540
Db	481	GGCCATGAAAGGTACTGATACGCTCAAGTCGCGGCTCGGTTACATCCTGCTTACTA	540
OY	541	CGTGTGTACTGTGGTCTCTGTCTGTATTAATTAAGTATGAGCTTCACGCCACTCTTCC	600
Db	541	CGTGTGTACTGTGGTCTCTGTCTGTATTAAGTATGAGCTTCACGCCACTCTTCC	600
OY	601	ATGGGCTATTTTGTACGCTGAGTGGAGAGAACTGACCCCTGATGATCCAAACCTTGTC	660
Db	601	ATGGGCTATTTTGTACGCTGAGTGGAGAGAACTGACCCCTGATGATCCAAACCTTGTC	660
OY	661	ATCAGTCACAAACATACCAATGTGTACAGCAGTGTCTAAGTCATCTTTTGAAGACAGT	720
Db	661	ATCAGTCACAAACATACCAATGTGTACAGCAGTGTCTAAGTCATCTTTTGAAGACAGT	720
OY	721	TCTCCAAACAAAGCGATGGAATGAAGGAGGTGCGGTGCGCCCATCTGTAATGTTGTT	780
Db	721	TCTCCAAACAAAGCGATGGAATGAAGGAGGTGCGGTGCGCCCATCTGTAATGTTGTT	780
OY	781	GTGTCTATTCATGCGATGGCTCATGTGTTGCGAGTGTGCGCCGAGAGAGTCMAAGATTG	840
Db	781	GTGTCTATTCATGCGATGGCTCATGTGTTGCGAGTGTGCGCCGAGAGAGTCMAAGATTG	840
OY	841	CGGCAAAAGGCGCTACTTCCTGCGGCTCTTCCCATACGTTGTGATGATCATTATTAT	900
Db	841	CGGCAAAAGGCGCTACTTCCTGCGGCTCTTCCCATACGTTGTGATGATCATTATTAT	900
OY	901	CACCAACAATCAATCCTGCGCGGTGTACGTACAGCGGATACCTGTTCTGTCAGCGCTCATG	960
Db	901	CACCAACAATCAATCCTGCGCGGTGTACGTACAGCGGATACCTGTTCTGTCAGCGCTCATG	960
OY	961	GGCGAAACCTCTGAGTGTGCGGTGTATGTACTCAGCACTCAGCAAGTGTCTCTCT	1020
Db	961	GGCGAAACCTCTGAGTGTGCGGTGTATGTACTCAGCACTCAGCAAGTGTCTCTCT	1020
OY	1021	GACAGTGTGACCGGACCGATCATCATCTTCTCCCTTACAAGGTTTCAACATTAATAT	1080
Db	1021	GACAGTGTGACCGGACCGATCATCATCTTCTCCCTTACAAGGTTTCAACATTAATAT	1080
OY	1081	CTACAGGATGCTTGATTTGTACGACTTGTGACACCTTTACAAGTTCCTGTCTGGGT	1140
Db	1081	CTACAGGATGCTTGATTTGTACGACTTGTGACACCTTTACAAGTTCCTGTCTGGGT	1140
OY	1141	CACGATCTGCGTATTCCTTGTATACCTGCGGTACAGACTCAACATCAACAGAGTGGAGATGT	1200
Db	1141	CACGATCTGCGTATTCCTTGTATACCTGCGGTACAGACTCAACATCAACAGAGTGGAGATGT	1200
OY	1201	GGTGTGTGTGCGGAGTACAGACTTGTGCTTTCATTTACCTGATGATGATGATGATGATGATG	1260
Db	1201	GGTGTGTGTGCGGAGTACAGACTTGTGCTTTCATTTACCTGATGATGATGATGATGATGATG	1260
OY	1261	ATTCCAAACCTCAGTATTTCTGGTGTCTTCTGATGATGATGATGATGATGATGATGATGATG	1320
Db	1261	ATTCCAAACCTCAGTATTTCTGGTGTCTTCTGATGATGATGATGATGATGATGATGATGATG	1320

Db 1261 ATTCCAACTGAGTATCTCGGCTGCTCTCTGATGATGTCGGTGGGTACG 1320  
 Qy 1321 CTATCCGCTGCTGCTATCGACTTCAACACATGGGAGCGCTTCCAGCTG 1380  
 Db 1321 CTATCCGCTGCTGCTATCGACTTCAACACATGGGAGCGCTTCCAGCTG 1380  
 Qy 1381 ACCACGCTACATGTCAGGATGACCTGTTCTGGCGTTCCTGCTGACTGTTA 1440  
 Db 1381 ACCACGCTACATGTCAGGATGACCTGTTCTGGCGTTCCTGCTGACTGTTA 1440  
 Qy 1441 CTGACACCGGCTGGACATATATCTTGACCTGTGATCAGAGGTAACATCT 1500  
 Db 1441 CTGACACCGGCTGGACATATATCTTGACCTGTGATCAGAGGTAACATCT 1500  
 Qy 1501 TGTGCTTTCTGCGCATTTCTGAACGCGAGCGGCTTCTGATTTATGATGAGAA 1560  
 Db 1501 TGTGCTTTCTGCGCATTTCTGAACGCGAGCGGCTTCTGATTTATGATGAGAA 1560  
 Qy 1561 TCTGCTCTAGACATTTGATTCATGTTGGTAAAAAGACTGCTTACGCTCTGCTC 1620  
 Db 1561 TCTGCTCTAGACATTTGATTCATGTTGGTAAAAAGACTGCTTACGCTCTGCTC 1620  
 Qy 1621 CTGGGCGCTAATCATCTCTGCTATATAGACGCTGTTCTTACGCTCTGCTC 1680  
 Db 1621 CTGGGCGCTAATCATCTCTGCTATATAGACGCTGTTCTTACGCTCTGCTC 1680  
 Qy 1681 TAACAACCTGCTGTCGAGACACTAGATATACCGACTGCTGTTATGTTCTGATA 1740  
 Db 1681 TAACAACCTGCTGTCGAGACACTAGATATACCGACTGCTGTTATGTTCTGATA 1740  
 Qy 1741 CTGTGATTTATTTTGGGCTACCTGTTGCTCAATTTGGAATTTGTTTGTACAA 1800  
 Db 1741 CTGTGATTTATTTTGGGCTACCTGTTGCTCAATTTGGAATTTGTTTGTACAA 1800  
 Qy 1801 ATACGCTACCGAATCTTACGAGACATCAAGAAAGCTTCCATCCAAACCTCATG 1860  
 Db 1801 ATACGCTACCGAATCTTACGAGACATCAAGAAAGCTTCCATCCAAACCTCATG 1860  
 Qy 1861 GGGTCCCGCTCGCGAGAGAGCGTGAAGATGATGCAAGCTGGAAGCGAAAC 1920  
 Db 1861 GGGTCCCGCTCGCGAGAGAGCGTGAAGATGATGCAAGCTGGAAGCGAAAC 1920  
 Qy 1921 TCTTACGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980  
 Db 1921 TCTTACGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980  
 Qy 1981 TTACAGCGCTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040  
 Db 1981 TTACAGCGCTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040  
 Qy 2041 CTACTGATCTTATGTTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
 Db 2041 CTACTGATCTTATGTTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
 Qy 2101 CAAAGGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 Db 2101 CAAAGGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 Qy 2161 TATGACGATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220  
 Db 2161 TATGACGATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220  
 Qy 2221 GTATATGTTTATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280  
 Db 2221 GTATATGTTTATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280  
 Qy 2281 AATAACATTTTATACATGTTGATGATGATGATGATGATGATGATGATGATG 2340  
 Db 2281 AATAACATTTTATACATGTTGATGATGATGATGATGATGATGATGATGATG 2340  
 Qy 2341 AAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400  
 Db 2341 AAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400

Qy 2401 CCATTTATCAGCTAGACCGCAAGATGATGATGATGATGATGATGATGATGATG 2460  
 Db 2401 CCATTTATCAGCTAGACCGCAAGATGATGATGATGATGATGATGATGATGATG 2460  
 Qy 2461 CCAATTTGCTAATAATCAAACTTCAATTAATTAATTAATTAATTAATTAATTAAT 2520  
 Db 2461 CCAATTTGCTAATAATCAAACTTCAATTAATTAATTAATTAATTAATTAATTAAT 2520  
 Qy 2521 TATACATTAACCGGCTGATATTAATTAATTAATTAATTAATTAATTAATTAAT 2580  
 Db 2521 TATACATTAACCGGCTGATATTAATTAATTAATTAATTAATTAATTAATTAAT 2580  
 Qy 2581 GTGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640  
 Db 2581 GTGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640  
 Qy 2641 ATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700  
 Db 2641 ATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700  
 Qy 2701 GTACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760  
 Db 2701 GTACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760  
 Qy 2761 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2820  
 Db 2761 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2820  
 Qy 2821 GATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880  
 Db 2821 GATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880

RESULT 2  
 AF006063 2881 bp mRNA linear INV 24-JUN-1998  
 LOCUS AF006063  
 DEFINITION Manduca sexta potassium coupled amino acid transporter (KAT1)  
 mRNA, complete cds.  
 ACCESSION AF006063  
 VERSION AF006063.1 GI:3252835  
 KEYWORDS  
 ORGANISM Manduca sexta.  
 SOURCE Manduca sexta.  
 REFERENCE 1 (bases 1 to 2881)  
 Hediger, W.A.  
 Cloning and characterization of a potassium-coupled amino acid transporter  
 Proc. Natl. Acad. Sci. U.S.A. 95 (9), 5395-5400 (1998)  
 2 (bases 1 to 2881)  
 Castagna, M., Shaykhal, C., Trotti, D., Sacchi, F.V., Harvey, W.R. and  
 Hediger, W.A.  
 Direct Submision  
 Submitted (30-MAY-1997) Renal Division, Department of Medicine,  
 Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115,  
 USA

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 1. .2881  
 /organism="Manduca sexta"  
 /db\_xref="taxon:7130"  
 1. .2881  
 /gene="KAT1"  
 103. .2007  
 /gene="KAT1"  
 /codon\_start=1  
 /product="potassium coupled amino acid transporter"  
 /protein\_id="AAC24190.1"

gene  
 CDS  
 103. .2007  
 /gene="KAT1"  
 /codon\_start=1  
 /product="potassium coupled amino acid transporter"  
 /protein\_id="AAC24190.1"

/db\_xref="gi:3252836"  
/translation="MNDGOVNGSESEPEKMEPRSSQISLPANIKATWINDIDML  
EAPPEMNSNTEPMSCTANSVGLGNWREPELAYONGGAFVLPYIVLLVKG  
PVYLEEVLOFSSRSNYSKIMWISPAKGGIQAQVCGGILSYVVICLFLYAM  
SFOATLWMAICOPENWCVSDPLTLAGVGNITNGSSABEYVIVLQSDIEBGL  
GAPLWYLVLCFLAMLVGVAVGVSNAKAYFLALPEYVYVITLFTYVILPAT  
DGLIEFTVPMWKLLEGLVMAITOVFEFLSVCTGATIMESYNOFNKVRDIALV  
TTLDTFTSFGGCIIEGILGNLAVELNSDGDVIGSGSISLISPDIAKTFDOLV  
FSVLEFLMSVGLGSSVALLSTLNVMDAPRPVYVMSALSCYGLLVYCTP  
GGPILVLDVHGSTPLVLCALSELNAGVWITGLENLCLDEFMKKTGGYWRLCW  
GIYVPMIIVFYALISLENLVFGDFVYVPAVYVAGYVIMLEFLVLPIDIVILY  
KYRGNRETVKAFHSKPSWGRPRRLREWMQFRIEAKALROKMNISRVKHWMSI  
TGAIRRNIN"

BASE COUNT 770 a 607 c 610 g 894 t

ORIGIN

Query Match 84.7% Score 2423.4; DB 3; Length 2881;  
Best Local Similarity 91.6%; Pred. No. 0;  
Matches 2627; Conservative 0; Mismatches 221; Indels 21; Gaps 5;

QY 9 GTTACTGTTGGAGAACTGTTGGCGGTGTCGATGTTCTGAACACAAATACCT 68  
DB 11 GTTACTGTTGGAGAACTGTTGGCGGTGTCGATGTTCTGAACACAAATACCT 70  
QY 69 AAACACATTGCAAGTGTGATTTGGACAAAATGATGACGCGCAAGTAAACGGGCT 128  
DB 71 AAACACATTGCAAGTGTGATTTGGACAAAATGATGACGCGCAAGTAAACGGGCT 130  
QY 129 TCGAGTCGTCGAGCCCAAGATGAAACCAAGATCATCGCAATTAAGTTACCTCAG 188  
DB 131 TCGAGTCGTCGAGCCCAAGATGAAACCAAGATCATCGCAATTAAGTTACCTCAG 190  
QY 189 CAATTAATTAAGCGGCTGTGATTAATTAATGATGACAGGACTTAGAGCTGACCGCAG 248  
DB 191 CTAATCTTAAGCGGCTGTGATTAATTAATGATGACAGGACTTAGAGCTGACCGCAG 250  
QY 249 AACGATGATGTCGCAACAACATTAATCTTATGTCGATGTCGATGTCGATGTCG 308  
DB 251 AACGATGATGTCGCAACAACATTAATCTTATGTCGATGTCGATGTCGATGTCG 310  
QY 309 GTTGGGTAAAGTGGGGGTCCTTTCATCGGTCACAGATGAGAGAGGCTTTC 368  
DB 311 GTTGGGTAAAGTGGGGGTCCTTTCATCGGTCACAGATGAGAGAGGCTTTC 370  
QY 369 TGTGTCATACGTCATGCTTCTTACTTGTGCGCAAGCCGTGCTACTTACTAGAGTGTG 428  
DB 371 TGTGTCATACGTCATGCTTCTTACTTGTGCGCAAGCCGTGCTACTTACTAGAGTGTG 430  
QY 429 TCTCGGCAATTAAGTTCAAGAACTGCTGTTAAAGTTGGTCAATTTACCGGCCATGA 488  
DB 431 TCTCGGCAATTAAGTTCAAGAACTGCTGTTAAAGTTGGTCAATTTACCGGCCATGA 490  
QY 489 AAGTACTGATGATGCTCAAGCTGCGGTCGCGGTTTCAATCGTCTTACTACTGAGTGA 548  
DB 491 AAGTACTGATGATGCTCAAGCTGCGGTCGCGGTTTCAATCGTCTTACTACTGAGTGA 550  
QY 549 TCTGTCGTCCTGCTCTATTAATTAAGTGAAGTCTTCAAGGCTTCTTCAATGAGGCTA 608  
DB 551 TCTGTCGTCCTGCTCTATTAATTAAGTGAAGTCTTCAAGGCTTCTTCAATGAGGCTA 610  
QY 609 TTTGTGACCTGAGTGGAGAACGCTGCTACCAATCCACTGCTGCTGCTGCTGCTGCTA 668  
DB 611 TTTGTGACCTGAGTGGAGAACGCTGCTACCAATCCACTGCTGCTGCTGCTGCTGCTA 670  
QY 669 ACAACATGCAATGATGACGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 728  
DB 671 ACAACATGCAATGATGACGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 730  
QY 729 AAGGCAATGCAATGATGAGAGGTCGCGGTCGCGGTCGCTGCTGCTGCTGCTGCTGCT 788  
DB 731 AAGGCAATGCAATGATGAGAGGTCGCGGTCGCGGTCGCTGCTGCTGCTGCTGCTGCT 790  
QY 789 TCAATGCAATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 848

DB 791 TCAATGCAATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850  
QY 849 CGGCTACTTCTGCGGCTCTTCCCATACGTTGATATATATATATATATATATATATATAT 908  
DB 851 CGGCTACTTCTGCGGCTCTTCCCATACGTTGATATATATATATATATATATATATATAT 910  
QY 909 TCAATGCAATGCTGATGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 968  
DB 911 TCAATGCAATGCTGATGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 970  
QY 969 TCAATGCAATGCTGATGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1028  
DB 971 TCAATGCAATGCTGATGAGTGTCAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1030  
QY 1029 GCACCGGCAAT 1088  
DB 1031 GCACCGGCAAT 1090  
QY 1089 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148  
DB 1091 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150  
QY 1149 TCGGTAATCTTGGTAACTGCGGTGACGAACACTCAAGTGGAGATGATGATGATGATGAT 1208  
DB 1151 TCGGTAATCTTGGTAACTGCGGTGACGAACACTCAAGTGGAGATGATGATGATGATGAT 1210  
QY 1209 CTGCGGCTACAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1268  
DB 1211 CTGCGGCTACAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1270  
QY 1269 CTGCGGCTACAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1328  
DB 1271 CTGCGGCTACAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1330  
QY 1329 TGGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1388  
DB 1331 TGGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390  
QY 1389 TCTCAATGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1448  
DB 1391 TCTCAATGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1450  
QY 1449 CGGCTGACAAAT 1508  
DB 1451 CGGCTGACAAAT 1510  
QY 1509 TCTGCGGCAATTTCTGACCTGCGAGGCTGCTGATTTATGATTTGAGATGATGATGATG 1568  
DB 1511 TCTGCGGCAATTTCTGACCTGCGAGGCTGCTGATTTATGATTTGAGATGATGATGATG 1570  
QY 1569 TAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1628  
DB 1571 TAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1630  
QY 1629 TAATGATGCTGATTAATTAAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1688  
DB 1631 TAATGATGCTGATTAATTAAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1690  
QY 1689 TGTGTTGGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1748  
DB 1691 TGTGTTGGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1750  
QY 1749 TATTTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1808  
DB 1751 TATTTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1810  
QY 1809 CCGGACATGCTGAGGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1868  
DB 1811 CCGGACATGCTGAGGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1870  
QY 1869 GCTGCGGAGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1928



```

Db      2322 TGGAAACAAGCATGAGTTCCTCATGTCCTCATGACGACGAGGCTGGGTGGCCAAAT 2381
Oy      320 GGTGGCGGTTCCTTCATCCGCGTACCAAGATGAGAGAGTGTCTTCTCGTGGCCATAC 379
Db      2382 GTGTGGAGGTTTCCCTTACCGCACTGGAGAACGAGAGAGGAGCCCTGTGTATACCCCTAC 2441
Oy      380 GTCATGCTTCTTTTACTGTGTGCGCAAGCCGTGTACTAGTGTAGTGTGTCTGTGGACAA 439
Db      2442 CTGATCGTCGTATCTAGTGGCAAGCCGCTGTACTGTGAATGATGCTGTGGCCAG 2501
Oy      440 TTCACTTCAAGAACTCTGTAAAGTTTGTCAATTTTCACGGCCATGAAAGTACTGA 499
Db      2502 TTCTGAGTTCGGGCGAGGTGAAAGGTCTACGACTTTTCAACCATCATGAGAGGCAATTGA 2561
Oy      500 TACGCTCAAGCTGCGGCGGTACATCCGTGTCTTACTAGTGTGTATCTGTGTCTC 559
Db      2562 TATGGAACAAGTCGTGGCCAGCGGCACTGTACACCTTACTAGCCACTTAATGGCGCTC 2621
Oy      560 TGTCTGTATCTAGTATAGCTTCCAGGCACTTTCATGTGGCTATTTGTACGCT 619
Db      2622 ACCCTACGCTACTTGTGAGACTCTTACCCGACCTTGCCTGTGAGCTACTGTGGCGAG 2681
Oy      620 GAGTGGG-----AGACTGGCTACCCCTCAGATCCAACACTTGTCTGATCAGTC 667
Db      2682 GAGTGGGTACAGAAATGCTGTGACTCGGGACCGGAGCGGACGACAGACCAACCAAGTTTG 2741
Oy      668 AACAAATCACCAATGTATACAGAGTGTCACTACTTTTGTAGAAAGATTTCTCCAA 727
Db      2742 GCAGGGTTCGGGAGTGGCAACCACTCGGCGGAGATTTCTATTCACGAACATATTTCTGCC 2801
Oy      728 CAAGCGATGGAATTTGAAGAGTCTGTGGTCCCCCATCTGTGTACTTGTGTGTCTA 787
Db      2802 GAGAGGGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2861
Oy      788 TTATCGCATGGCTCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 847
Db      2862 GCTGTGGCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2921
Oy      848 GCGGCTTATCTCTGCGGCTTCCATACGTTGTCAATGATCAATTTATCATCAGACA 907
Db      2922 GCGCTTACTTCTCTGCGGCTTCCATACGTTGTCAATGATCAATTTATCATCAGACA 967
Oy      908 ATCATCTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 967
Db      2982 CTCACCTTGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3041
Oy      968 CTCCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1027
Db      3042 CTCCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3101
Oy      1028 TGCACCGGACCATCATCATGTTCTCTTACCAAGGTTTTCAGACATTAATATCTACAG 1087
Db      3102 TGCCTTGGGAACATCATCATGTTCTCTTACCAAGGTTTTCAGACATTAATATCTACAG 3161
Oy      1088 GATGCTTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1147
Db      3162 GATGCTTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3221
Oy      1148 TTGGATTCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1207
Db      3222 TTGGATTCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3281
Oy      1208 GCTGGGTTCAGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1264
Db      3282 AATGGAGGTTCAGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3341
Oy      1265 CAACCTCATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1324
Db      3342 CTGCGGAGATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3401
Oy      1325 TCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1384

```

```

Db      3402 AATGGGCAATGCGCTCTGATGTCCAGCTGATCAAGATCAAGTTCGACACCTGAG 3461
Oy      1385 ACCGTACATGATGACGATGACCTGTCTTGGCGTTTCCGTGTGACTGTACTGAC 1444
Db      3462 AACTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3521
Oy      1445 ACACCGGTGCAATATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1504
Db      3522 ACACCGGTGCAATATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3581
Oy      1505 CTTTTCTGCGCCATTTTCAAGCTGTGCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1564
Db      3582 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3641
Oy      1565 TGCCTAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1624
Db      3642 TGCGGTACCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3701
Oy      1625 GCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1684
Db      3702 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3761
Oy      1685 AACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1723
Db      3762 CCCCTCAAGTACAGAGATTTACCTTACCAATTCGGGTCTT 3800

RESULT 4
AX191697 2473 bp DNA linear PAT 15-AUG-2001
LOCUS AX191697
DEFINITION Sequence 3 from Patent WO0149848.
ACCESSION AX191697
VERSION AX191697.1 GI:15209877
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2473)
AUTHORS
Kellerman, K.A., Keegan, K.P., Ebens, A.J. and Torpey, J.
Nucleic acids and polypeptides of drosophila melanogaster snf sod1
sodium-neurotransmitter symporter family cell surface receptors and
methods of use
Patent: WO 0149848-A 3 12-JUL-2001.
JOURNAL
Genoptera, LLC (US)
FEATURES
Location/Qualifiers
Source
1..2473
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 625 a 593 c 627 g 628 t
ORIGIN
Query Match 13.7%; Score 391.8; DB 6; Length 2473;
Best Local Similarity 53.2%; Pred. No. 8.6e-75;
Matches 951; Conservative 0; Mismatches 767; Indels 69; Gaps 3;

```



QY 458 GTTAAAGTTGGTCAATTTACCGGCATGAAAGTACTGATACGCTCAAGCTGCCGC 517  
 DB 567 GTAAAGATTGGTGGTGGTGGGATTTGTGGCGTGGGATAGGCGAGGCTTCGGC 626  
 QY 518 TGGGGTACATCTGCTTACTAGTGTGATCTGTGCTGTCTGTACTTACTTACT 577  
 DB 627 AACATCTGCATCTCTGCTACTTCTCTGCTGTGGCGCTGACCCCTCTACTCTCTT 686  
 QY 578 ATAGAGTTCAGGCGCTCTTCCATGAGGCTATTGTTCAGGCTGAGTGGAGAACTGGCTA 637  
 DB 687 GTGTCTTCCATTCGAAATGCGGTGGTCTACTGTGCGCATGAGTGGACCAATTGGCTG 746  
 QY 638 CCCGC----- 642  
 DB 747 AACTCGAGCCGACGAGTATGTGACAAATCTGCTTACGGGCGTATCTTGGCCATGAA 806  
 QY 643 -----AGATCCAACTTGTCTGATCAGTCAACAACTCAACAAATGGTACAGAGT 694  
 DB 807 TCGGCCAAGATCTCAGCGGCGATTTGTGCCACAGATAGACAGAAACTCCAGACAGC 866  
 QY 695 GCTCACTACTTCTTTGAGAACAGTCTCCACAAAGCGATGGAATTGAAGAGTCTC 754  
 DB 867 TCCGAACTTATTTCTTGAATGTAGTATCAGAGAGAGAGCTGACATCTCGAGCGGCTT 926  
 QY 755 GGTGCCCCCAGTGGTACTTGGTGTGTATTCATCGCATGGCTCATGCTGTGGGA 814  
 DB 927 GGTATCCCGATTGGAAGCTGACACTTGCCTGTTCGCTGCGCTGGTGGTATCTTTCTG 986  
 QY 815 GTGCTGCCGAGAGAGTCCAGAGTTCGGGCAAGCGGCTTACTCTCGCGCTTCCCA 874  
 DB 987 GTGATCATGCGAGGGGTGAAGAGTTCGGGCAAGCGGCTTACTCTCGCGCTTCCCA 1046  
 QY 875 TACCTGTCAATGATCACTTTATTCATACCAACCAATCAATCCCGCGGCTACTAGAGCGC 934  
 DB 1047 TACGTGTGCTCTTCTGTCTGTCTGATCCGCTTACTTGAAGAGAGCAGCGAGTGC 1106  
 QY 935 ATCTGTCTCTGTCAGCGCTCAATGAGGCGCAAACTCTTGAAGTGTGTGATCTCA 994  
 DB 1107 ATTCTGTCTTCTGAGGCGCAAGTGGGTGATCACTTGAACCCAGCTTTGGAAGAG 1166  
 QY 995 GCAGTCAAGCAAGTGTCTCTCTGACAGTGTGACCGGACCGCATCATGCTTCTCC 1054  
 DB 1167 GCCCTGTGAGTCTCTCTCTGCTGCGCTGGGCTCCGAGCCCATATCATGCTTCC 1226  
 QY 1055 TCTTCAACAGGTTTCAAGCATATATATACAGGAGTGTGATGTTAGCACTTGGAC 1114  
 DB 1227 TCGTACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1286  
 QY 1115 ACCTTACAGATTTCTGTCTGAGTGCAGATCTTCTGATCTTGTGATACCTCGCTGAC 1174  
 DB 1287 ACGGTGACACCTCTCTGAGAGTATACGATATTTGGATCTTGGGCAATCTGGCGAC 1346  
 QY 1175 GAAGTCACTCAGAGTGGAGATGTGTGCTGCTGCGGTACCAGTCTTGTCTTCAAT 1234  
 DB 1347 AATCTGAGATCGAAACATCAGAGATGTGTGCGAATGCGACGAGTGTGGCTTCAATC 1406  
 QY 1235 TCATACCTGATTCGATTCGCA---AATCTTCAACCTCACTATTCGCTGCTGCTG 1291  
 DB 1407 TCGTATCCGATTCGATTCGCAAGTTCAGGCGGTGCGCGACCTTTTCTCGCTTCTGCTC 1466  
 QY 1292 TTCTGATGATGTGCTGCTGCTGATTCGCTATCCGCTGCTGCTGCTGCTGCTGCTGCT 1351  
 DB 1467 TTCTTCAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1526  
 QY 1352 ACATTTGGGAGTGGAGCGCTTCCACAGCTTACCACCGCTTATCTGACGATGACCTGT 1411  
 DB 1527 AACATCATTTGCGATTCAGAGGATGGAAGTACTGGAAGTGGCAGCTGACCACTCT 1586  
 QY 1412 TCTTGGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1471  
 DB 1587 GTGTGCGGTTCTTATGAGGTTTGTCTACGTTACCGGCGGAGAGTGGATCTCTCACT 1646  
 QY 1472 CTGTGATCACTACGAGTGAACATCTCTGTGCTTTCTGTGCGCATTTCTGAACTCGCA 1531

DB 1647 TTGGTGAATCTTATGCGGACCATATGCTGTCTTACTTCTGGCCATTTTGCAGAACTGCT 1706  
 QY 1532 GCGGTGTTCTGATTTATGATTTGAGATCTGTGCTTACATTTGATTCATGTTGGGT 1591  
 DB 1707 GGCATGCTGTGGTATGCTGTCGAAAACTTTTGGATGATCATGCGATTTATGTCAT 1766  
 QY 1592 AAAAAGCTGTGCTTACTGCGCTCTGCTGGGCGGTATTCATCTCTGATTAAGCAG 1651  
 DB 1767 CGCGAGTTTCACTGATCTGCGAGTATGCTGGTCTTCTTCAAGCGCCGATGATGATC 1826  
 QY 1652 ACTGTGTTCTTCAAGCTCTTCTGCTTCAACAACTGCTGCTGAGACAACTACGTA 1711  
 DB 1827 ATCATATTCAATATCTGATGCTGATGCTGATGATGATGATGATGATGATGATGATGAT 1883  
 QY 1712 TACCGAGCTGCTGATTTGTTCTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1771  
 DB 1884 TTCCCGAAGCGGCAACATTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943  
 QY 1772 CCATTTGGAATTTGATTTTCTTGTACAAATACCGTACCGGAACTCTGAGAGAGATC 1831  
 DB 1944 CCATGTTGGGCTTTTGTGATCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2003  
 QY 1832 AAGAAAGCTTCCACTCCAAACCTTCAATGAGGCTCCGCTGCGGAGAGAGAGAGAG 1891  
 DB 2004 AAGGCTGCGGAG 2063  
 QY 1892 TGATGCACTTCAAGGCTGGAAGCAAGCTTTGACAAAAGATGAA 1938  
 DB 2064 TGGGTGATCTTCAAGATCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110

## RESULT 5

LOCUS AY121662 2497 bp mRNA linear INV 18-JUN-2002  
 DEFINITION Drosophila melanogaster RE10560 full insert cDNA.  
 ACCESSION AY121662.1 GI:21464371  
 VERSION AY121662.1  
 KEYWORDS FLI.CDNA.  
 SOURCE fruit fly  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 2497)  
 Stapleton, M., Brokstein, P., Hong, L., Aghayani, A., Carlson, J.,  
 Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,  
 George, R., Gonzalez, M., Guarin, H., Krommiller, B., Li, P., Liao, G.,  
 Miranda, A., Mungall, C. J., Nunoo, J., Paclebb, J., Paragas, V., Park, S.,  
 Patel, S., Phouenavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M.  
 and Celinker, S.

## TITLE

Direct Submission

## JOURNAL

Submitted (13-JUN-2002) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory

## COMMENT

This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unsplined precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our Web site  
 (<http://fruitfly.berkeley.edu>) or send email to

```

cdnaefruitfly.berkeley.edu.
Location/Qualifiers
1..2497
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db.xref="taxon:7227"
1..2497
/gene="CG3252"
/db.xref="FLYBASE:FBgn0029762"
280..2205
/gene="CG3252"
/note="Longest ORF"
/codon_start=1
/product="REL10560p"
/protein_id="AA051989.1"
/db.xref="GI:2146372"
/db.xref="FLYBASE:FBgn0029762"
/translation="MELKGVQPSNGSSNGSNGATNAASTEKTDKPTAERTNMGNG
LEFLMSTISVSVGLGNWREPTAYENGGAFLPIYILVFLGKPMYLEMIMGOT
SOGTKTMSVYVPEFVGQYGAQFETICIIYSYSLALTYLYLFSQSELPMSCPD
EMTNCVNSRQEVNDLITGSLANESARLISGVANDEKLOSSELYFLANVYKE
KLIDSDGVDPDKLTLALFVAVVYFLVMRGYKSSKRAYFLAIPRYVLELIR
AVTLEGARDGILFLEPQWGEILNPTVKRAVVOCPSLAVSGSPITMPASINREBHG
IYRDAMVITLDLITSLGSIITPAILIGNAHNIQENIDVRSGLAFISYPAI
SKRDVQDLSVLEFFMFLVGLGSLVALDSTLVITICDQPKMKRWKVALTYSVCF
LMGLVYTPGGQWILTLVDFYGGYVFLAIFELAGIYVWVIGLONCDIEMCNBR
VSLRWVMSRFPVMMIIFISYMWIIEIKYSELFEPAANAGLWALFAICAOFP
LMGLMYISRHQCTYMKSLKSLKPSDRMGPAPEIRREVIIFKNQAAQRAQKNGS
KLGFPMKRVNFCGSNK"
BASE COUNT      625 a      604 c      636 g      632 t
ORIGIN
Query Match      13.7%; Score 391.8; DB 3; Length 2497;
Best Local Similarity 53.2%; Pred. No. 8.6e-75;
Matches 951; Conservative 0; Mismatches 767; Indels 69; gaps 3;
QY 218 GATGACGCGACTAGAGCGTGAACCGCCAGAACGTATGATGTCACAAACATGAA 277
Db 358 GAAAAACACATGCCGGAAGCCGCGGAGCGCACCAATTGGGCAATGGCCTGGAG 417
QY 278 TTCTGATGTCGTGACATCGCTACATCCGCTGGTGGTGAAGCGTGGCGGTTCCCTTC 337
Db 418 TTCTATATGTCGTGACATCGGTGTCGTTGGGANTGGGCAACGCTGGAGGTTCCCGTTC 477
QY 338 ATGCGGTACCAAGATGAGAGAGTCTTTCTGCTGTCATACGTATGCTTTTACTT 397
Db 478 ACGGCTATGAGATGAGAGAGCGGCTTCCTTATACCTACATCATAGTCTCTTCTG 537
QY 398 GTGCGCAAGCGCTGTACTACTAGATGTCCTCGGACATTCAGTTCAANAACCTC 457
Db 538 ATGGGCAACCCATGTACTATTGGAGATGATCATGGGCCAGTTCAAGAGCCAGGGAACG 597
QY 458 GTTAAAGTTGGTCAATTTCACGGCCATGAAGATGATGATACGCTCAAGCTGCGGC 517
Db 598 GTCAAAATTTGGTCGTGGTGGCGGATTTGTGGCGGTGGCGATGAGCCAGGCGCTTCGG 657
QY 518 TGGGTTACATCTGTCCTTACTACGCGGTGATCTGCGTCTGCTGTAATTAAGT 577
Db 658 ACCATCTGCATCATCTGTAATTCCTCGCTGTTGGCGGTGACCCCTCAATCTCTTT 717
QY 578 ATGAGCTTCAGGCGCACTCTTCATGAGCGGATTTGTCAGCTGAGGGAGGAACGCTA 637
Db 718 GTGTCTTCCAAATCGGAATTTGCCGTGCTTACTGTGCGGATGAGTGAACCAATTCG 777
QY 638 CCCTC----- 642
Db 778 AACTCGAGCCGAGAGATATGACAATCTGCTTACGGCGGTATCCTTGGCCAATGA 837
QY 643 -----AGATCAACACTTGTGATAGTCAACACATCATCAACATGTAACGAGAT 694
Db 838 TGGGCCAGAAATCTCAGCGCATTTGTGGCAACGATGAGACAGAGAACTCCAGACAC 897
QY 695 GCTCAACTCTACTTTTGGAAACAGTTTCCAAACAAGGATGATGAAGAGAGCTGC 754

```

```

Db 898 TCCGAACCTCTATTTCCTGAAATGATGATGATGAAGAGAACTGGACATCTCGAGCGGCTT 957
QY 755 GGTGCCCACTGCTGACTTGGTGTGTCTATTCATCCAGATGCTCATGTGTTCGGA 814
Db 958 GGTGATCCCGATTGGAAGCTGACACTTGCCCTGTGCTGCGCGGGTGTGATCTTTCTG 1017
QY 815 GTCTGCCCGGAGAGATCAAGAGTCTCGGCAAGCGGCTTACTTCTCGCGCTTCCCA 874
Db 1018 GTGATCAATGAGAGGAGTGAAGAGTCCGCAAGCGGCTTACTTCTCGCGCTTCC 1077
QY 875 TACGTTGTCATGATCACTTTATTCATCACAATCATCTCGCCGCTGCTACTGAGCG 934
Db 1078 TACGTGTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
QY 935 ATCTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 994
Db 1138 ATCTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
QY 995 GCAGTCACGAAGTCTCTCTCTGACAGTGTGACAGCGGACCGATCATATGTTCTC 1054
Db 1198 GCGGTGTGAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
QY 1055 TCTTACACGCTTTCAGACATTAATATCTACAGAGATGCTGATGTTGACATTTGAC 1114
Db 1258 TCGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1317
QY 1115 ACCTTTACAAAGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1174
Db 1318 ACGCTACACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
QY 1175 GAATCAACATCAGAGTGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1234
Db 1378 AATCTGCAATGCAAAACATCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1437
QY 1235 TCAATACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291
Db 1438 TCGTATCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497
QY 1292 TTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1351
Db 1498 TTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
QY 1352 ACATTTGCGATGAGCGCTTCCACGCTGTACCCAGCTGTACATGTACAGAGATGCT 1411
Db 1558 ACCATCATTTGCGATGAGCGCTTCCACGCTGTACCCAGCTGTACATGTACAGAG 1617
QY 1412 TCTTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1471
Db 1618 GTGTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
QY 1472 CTTTGTGATCAGAGTGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1531
Db 1678 TTGTGTGATCAGAGTGGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
QY 1532 GCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1591
Db 1738 GCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1797
QY 1592 AAAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1651
Db 1798 GCGCGAGTTTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
QY 1652 ACTGTGCTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1711
Db 1858 ATCATATTTATATATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1914
QY 1712 TACCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1771
Db 1915 TTCCCGGAAGCGCCCAACATTTCCGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1974
QY 1772 CCAATTTGAATTTGATTTTCTTGTACAAATACCTGCTGCTGCTGCTGCTGCTGCTGCT 1831

```





QY	1453	TGACAAATATATCTTGAGCTGTGAGATACATACGGGAGAACTCTGTGGCTTTCTG	1512
Db	1432	CGAGCACTTGCTACTGAATCTCGTCGACTTTTATAGCTGCACCTTTTATAGCCCTGGCTTT	1491
QY	1513	CGCCATTTTGCAGCTCGAGGCGCTGTCTGGATTTATGGATTGGAGATCTGTCCATGA	1572
Db	1492	GGCCATTGCAGAAATTTGTTGGCAGTGGGCTGGATTTATAGTGTGAAAGGCATCTGCAGGA	1551
QY	1573	CATTGAATTCAATGTTGGGTTAAAAAGACTGTGCTTACTGGCGTCTCTGTCTGGGCGTAAAT	1632
Db	1552	CATTGAATTCATCTGTCGATGTGAAGACAGCTTATTTGGCGCATCTGTGGGCAATTTGT	1611
QY	1633	CACCTCCGATTAATAGAGACTGTGATCTTCTTACGCTCTTCTCCCTCTACACCTGGT	1692
Db	1612	GGCACTCTGCTGTATGTTCTCGTGCTCTGCTATATGCTCTTCACTACGACACTCTGCAC	1671
QY	1693	GTTCGGAGACAACATCTAGTATACCCGACTGCTGTTATGTTCTTGAGTACTTGATTTATT	1752
Db	1672	GTACCGAGGCTTCA---GTATCGCGCGGTCTACTATATAGCTGGAATGGATTTATCTGGCG	1728
QY	1753	TTTGGGCAATGACGTTTGTGCCAATTTGGAATTTGATTTCTTTGACAAATATACGTAACGG	1812
Db	1729	ACTGTGGTGTCTGAGATTTGCCCTTCTGGGCTTTTGGTGCAGATTTACAGCGACGACGAGAA	1788
QY	1813	AACCTTAGCAGACGATCAAGAAAGGCTTCCATCCAAACCCGATCGAGGGTTC	1866
Db	1789	GACTTTTGGCAGTAAATTCGGACTGGCCATTCCAACCACTCCCAATTTGGGGTCC	1842

RESULT 8			
AY058396			
LOCUS	AY058396	2097 bp	mRNA linear
DEFINITION	<i>Drosophila melanogaster</i> Grl6l61 full length cDNA.		
			INV 16-OCT-2001

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.

**TITLE** Direct Submission  
**JOURNAL** Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,

COMMENT: Sequence submitted by: Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from *Drosophila* gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES	Location/Qualifiers
source	1. .2097

gene

CD5

BASE COUNT 522 a 498 c 514 g 563 t

ORIGIN

Query Match 13.2%; Score 378.6; DB 3; Length 2097;  
Best Local Similarity 53.2%; Pred. No. 6.4e-72;  
Matches 976; Conservative 0; Mismatches 799; Indels 58; Gaps 6;

/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="35B7-35B7"  
1..2097  
/gene="Bg:DS03431.1"  
/note="alignment with genomic scaffold AE003645"  
/db\_xref="FlyBASE:FBgn0028886"  
39..1958  
/gene="Bg:DS03431.1"  
/note="Longest ORF"  
/codon\_start=1  
/product="Gh16161p"  
/protein\_id="AAL13625.1"  
/db\_xref="GI:16183097"  
/db\_xref="FlyBASE:FBgn0028886"  
/translation="MDTTRRNHQLAEVGDGGRSTAVEISTNSPALRDNSDDEAA  
KYPERRATWKGVEFLFMSCLMSVGLANVMFPTALDNGGAFILPLVILFIFGR  
IYLEMYIAGPSSRSQSVAFEDLCPMKGVGQAGQAFQVEMLETTYAAALVILGRSE  
ERNLPMSTCKAENGHICINSAPDASNMQSESDORONTMKSNDORVITSEMEF  
VKELHKEPRNEEGTIGLPMNELVIGLPIAMSCVEFTIRRGVSSGKSAFLALPPVY  
MGVLVRAVTLPGSIDGIIYFIRKPQMKILDPKRYAAVYOCFSLSVCFNITMYSS  
FNKGHNHRAALVITGLDITSLAGTIFGLIHLHALEGTDDIGSVAGGAGGL  
ISYDAIAKFNELPQIFSVLEFPMLEFVIGISNSMCSVTALRDREPNRQMGSL  
LIAVVSFGLMAYITPGQGYMLTIVDFGSMIALVLVIGLAELTYIGMYGDRCKD  
EEMGRKGVGLWRLCMSITPLMIVLTIYATYOPLTNYIYPMNSYSIGMLIRKPL  
FGILQPLMIVLVAIVRDGQTLGAKIRGAFIPKKNMGSDPLANEQYHKEIENELTP  
RGQGIWMAIKONITG"

OY	88	TATTGTGGACAAATGATGTAGCGGCACAGTAAACGGCGGTTTCGAGTGTGGAGGCCAA	147
OY	27	TACACTGGCAATATGTGATATGACAGCGAAAGGAACCAACCAAAATCTAGCCTTGTGGGGCAC	86
OY	148	GATGAAACCAAAACGATCATCGCAAAATAGTTTACCTCCAGCAAAATAATTAAGGGGGCTCT	207
Db	87	GACGGACCCAGATACGGCCAGACAGTGGAGATTTGACCC---AACACTCCAGCCCTTCG	142
OY	208	GGATTATATATGATGTACAGGACTTAGAGGCTGAACCGGACGAGAGCTATGTATGGTCCAA	267
Db	143	AGATATATTCGAGTGCACAGAGGCGCCAAAGTGGCAGAGAGCGTCCACCTGGGGCCA	202
OY	268	CAACATTAATTCCTTGATGTGCTCGTCGATCCGTACATCCGTCGGTTTGGGTAAAGTGTGCG	327
Db	203	GGCGGTGAGATTTCATGTCTCGTCGATGCCCATGTCCGTTGGGTGGGAAAGTGTGGCG	262
OY	328	GTTCCCTTTCATCGCGTACAGAAATGAGAGAGGTCTTTCTGTGGTCCATATAGCTATGT	387
Db	263	ATTCGCCCTTACTCGCCCTCGATAAAGGGGGGGCCCTTCATACCATATATCTATTGT	322
OY	388	TCTTTACTTGTGCGCAAGCCGTGTACTACTTAAGTGTGCTCCGACAATTCATTTC	447
Db	323	CCTGTCTCTGATGGCAACCGCATTTACTATGTGGAATGGTAATTGGCCAGTTTTCAG	382
OY	448	AAGAAACTCTGTTAAAGTTTGGTCAATTTACCGGSCAATGAAGGTACTGATACGCTCA	507
Db	383	TGCGGGGCTCGGTTAAAGTGTTCGATTTGTGGCCGGAGTAAGAAAGTGTGGAGCGGCCA	442
OY	508	AGCTCCGGCTCGGTTACATCCCTGTCTACTACGTGTGGATGCTGTGGTCTGTGCTGTA	567
Db	443	GGCTTTCCAGGTTCATGCTAGCAGCATTAATATCCCGCCCTGGTAGCGATATATGGCCG	502
OY	568	TTACTTACCTATAGAGCTTCACAGGCCACTTTCATATGAGGGCTATTTGTACAGCCTGAGTGGGA	627
Db	503	ATATTTCATATGAATCATTTTGGCAACCCATTTGCCCTGGGTGCACATATGCCCGTGGGAATGGGG	562
OY	628	GA---ACTGCGTACCCCTCAGATC-----	647

Dh 563 TATCCACTGATTAACAGCCCCGATCGAGTAATGTCTCACTGGAAGCAATGA 622  
Qy 648 -CAACACTTGTGATCAGTCAACAATCAACAA-----TGTACACAGACTGCTCA 699  
Dh 623 TCAGCGCCCGCAAAATTTACACAATGAAATACAAAGAGATCGGTCTACACAGTTGCGA 682  
Qy 700 ACTCTACTTTTGGAAACAGTTCTCAACAAGCAGTGAATTTAGAGAGGTCTCGGTGC 759  
Dh 683 ATGCTACTTTTGAAGAAAGTGTGACAGAAACCAAAATATGAGGAAGAAATCGGCCT 742  
Qy 760 CCCCATCTGTACTGTGTGTGTCTATTCATTCAGCATGGCTCATGGTTTGGAGTGT 819  
Dh 743 ACCAAATGGGAGTGTGATTCATTTTATGATTCCTGCTGCTGTGTGTTCATTAAT 802  
Qy 820 CGCCCGAGAGTCAAGAGTTCGCGCAAGCGGCTACTTCTCGCGCTCCCATACGT 879  
Dh 803 TCGTCGGGAGTGAAGAGTTGGGAAAGCATCTACTTCTGCGCTTTTCCGTCAGT 862  
Qy 880 TGTCTATGATCATTATTCATCACCAATCATCTCCCGGTGCTACTGACGAGCATCT 939  
Dh 863 CATCATGGGTGTCTGTGTGTGAGCTGTAACTCTCGCGGATCAATAGCGTATATA 922  
Qy 940 GTTCTGTACGCGCTCAAGGGGCAACCTCTGAGCTGGGTATAGTACAGTACT 999  
Dh 923 CTACTTTATAAGCCGAATGGGAAGATTTTGGACCCGAAGTTTGGTACGCGGCTGT 982  
Qy 1000 CACGCAAGTGTCTCTCTGACAGTGTGACCGGACCGCATCATCTGTCTCTCTTA 1059  
Dh 983 AACTCAGTGTGTCTCTCTGACAGTGTGACCGGACCGCATCATCTGTCTCTCTTA 1042  
Qy 1060 CAACGGTTTCAAGATTAATCTACAGAGGATGCTTGGATTTTACAGCTTTGGACACTT 1119  
Dh 1043 CAACAAGTTGGGACATGATGACAGAGGATGACAGCAATTTGTAACGGGTGTGACACTAT 1102  
Qy 1120 TACAAGTTTCTGTGTGTGTGACAGATCTTGGATCTTGGTATCTGACAGTACT 1179  
Dh 1103 GACCTCTGTCTGTGTGTGATTCAGATATTCGATCTTGGCCATCTGGCCCAAGAAAT 1162  
Qy 1180 CAACCTCAGAGTGGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1239  
Dh 1163 CGGAACCGATGACATCGCTCAGTTGTGAAGGGGCGGTGGGATTTGGCTTCAATCATA 1222  
Qy 1240 CCCGATGACATTTGCCAA---AACATTCACAACCTCAGCTATTTCTGGTGTGTCTCT 1296  
Dh 1223 TCCCGATGACATTTGCCAA---AACATTCACAACCTCAGCTATTTCTGGTGTGTCTCT 1282  
Qy 1297 GATGATGTCTGT 1356  
Dh 1283 CATGCTGT 1342  
Qy 1357 GCGGATGAGACGCTTCCACAGCTTACCACCGGTCTACAGTGTGTGTGTGTGTGTGT 1416  
Dh 1343 CATTCGGGATCGGTTTCCCAACTTTGGGAGTGGGAGTGTGTGTGTGTGTGTGTGT 1402  
Qy 1417 CGGTTTCTGT 1476  
Dh 1403 CAATTTCTTTATTTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1462  
Qy 1477 AGATCACTAGAGTGAACATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1536  
Dh 1463 GGACTTTTGGGCGCTCAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1522  
Qy 1537 GTTCTGTGATTTATGAGATTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTT 1596  
Dh 1523 TGGATGATTTTAAAGGACAGATGCGCTGTGCAAGGAGATTAAGTGTGTGTGTGTG 1582  
Qy 1597 GACTGT 1656  
Dh 1583 AGTGGGCTGT 1642  
Qy 1657 GTTCTGTGAGCTTCTGT 1716  
Dh 1643 TCTGATCTA---TTTCTATGCGAGCTACCAACCTCTAACAATCAATATCTGTATCC 1699

Qy 1717 GACTGT 1776  
Dh 1700 AAATTTGTCTTACTCAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1759  
Qy 1777 TGAATTTGGAATTTTCTTTGTATCAAAATATACCGTACCGGAACCTTCAGGAGACATCAAGAA 1836  
Dh 1760 TTGATATATGT 1819  
Qy 1837 AGCTTCCACTCCAAACCTTCATGAGGTTCCCG 1869  
Dh 1820 AGCCTTACGCCGCAAAAGAAATTTGGGAGCCAG 1852  
  
RESULT 9  
AX191703  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
BASE COUNT  
ORIGIN  
Query Match 13.2% Score 378.6; DB 6; Length 2108;  
Best Local Similarity 53.2% Pred. No. 6, 4e-72;  
Matches 976; Conservative 0; Mismatches 799; Indels 58; Gaps 6;  
Qy 88 TATTGTGCAAAATATATGACGCGCAAGTGAACGCGGTTTGTGAGTGTGAGGCCAA 147  
Dh 38 TACAGTGGCAATTTGATATGACGACCAAGGAACCAAAATCTTCTGTGGGCGAC 97  
Qy 148 GATGGAACCAAAAGATCATTCGCAATTAAGTTTACCTCGAGCAAAATTAAGCGGCTCT 207  
Dh 98 GACGAGACGAGTATCGGCGACAGATGAGATTTGACG---AACAGTCCAGGCCCTGG 153  
Qy 208 GATATATATATATGACAGGACTTATAGGCTGACCGCCAGAAACGATATGTGTGCAA 267  
Dh 154 AGATTAATTCGATGACAGGAGGACCAAGGTGACAGAGGCGTGCACCTGGGGCAA 213  
Qy 268 CAACATTAATCTTGTATCTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 327  
Dh 214 GGGCGTGGAGTTTGTATATTCCTGATTCGATTCGATTCGATTCGATTCGATTCGATTC 273  
Qy 328 GTTCCCTTATTCGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 387  
Dh 274 ATTCCCTTATTCGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 333  
Qy 388 TCTTTACTTGTGTGCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 447  
Dh 334 CCGTTCTGATTTGGGCAACCGATTTACTGTGCAAAATGATTAATTTGGCAGTTTTCAG 393  
Qy 448 AAGAATCTGTATTAAGTTTGTGCAATTTCAACGCGGACGAAGGATGATGATGATGATG 507  
Dh 394 TCGGGGCTGTGTATTAAGTTTGTGATTTGTGCGCGGAGTGAAGGATTTGGAGCTGGCCA 453  
Qy 508 ACCTGCGGCTGT 567



```

Db 454 GCCTTCCAGGTGTCATGCTCAAGCACTTACTTACCGCCCTGGTACGCTAATTTGGCCG 513
QY 568 TTACTAGTATGATAGCTTCCAGGCCACTCTTCCATGGGCTATTTGTACGCTTGAAGGGA 627
Db 514 ATATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 573
QY 628 GA---ACTGGGTACCTTCAGATC----- 647
Db 574 TATCCAGTGTATTAACCTACGCCCCCGAGTGAATTTGTCTCACTGGAAGCAATGA 633
QY 648 -CAACACTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 699
Db 634 TCAGCGCCCGCAAAATTTACAAATGATGATGATGATGATGATGATGATGATGATGATGATG 693
QY 700 ACTCTACTTTTGGAAAGTCTTCCAAACAGGATGATGATGATGATGATGATGATGATGATG 759
Db 694 ATGGTACTTTGTTAAGGAAGTGTGACAGAGAAACAAATATCGAGGAAGGAATCGGCGCT 753
QY 760 CCCCATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 819
Db 754 ACCAAAGTGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 813
QY 820 CGCCGAGGAGTCAAGAGTTCGGCAAGCGGCTTACTTCCCTCGGCTTCCCATAGCT 879
Db 814 TCGTGGCGAGTGAAGAGTTCGGCAAGCGGCTTACTTCCCTCGGCTTCCCATAGCT 873
QY 880 TGTCAATGATCATTTTATTCATCACCACAAATATCTCTGCGGCTGTACTGACGCAATCT 939
Db 874 CATCATGGGTGTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933
QY 940 GTTCTGCTCAAGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 999
Db 934 CTACTTTTAAAGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 993
QY 1000 CACGCAAGTGTCTTCTCTGACAGTGTGACAGCGGACCAATCATATGTTCTCTCTTA 1059
Db 994 AACTCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1053
QY 1060 CAAGCTTTCAGACATTAATCTAGAGGATGCTTGGATGATGATGATGATGATGATGATGATG 1119
Db 1054 CAACAGATTTGGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1113
QY 1120 TACAGATTTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1179
Db 1114 GACCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1173
QY 1180 CAATCAGAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1239
Db 1174 CGGAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1233
QY 1240 CCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1296
Db 1234 TCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1293
QY 1297 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1356
Db 1294 CATGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1353
QY 1357 GCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1416
Db 1354 CATTCGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1413
QY 1417 CGGTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1476
Db 1414 CAGTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1473
QY 1477 AGATCACTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1536
Db 1474 GGAATTTTTCGCGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1533
QY 1537 GTTCTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1596

```

```

Db 1534 TGGATGATTTATGACAGATGCGCTGTGCAAGACATGATGATGATGATGATGATGATG 1593
QY 1597 GACTGGTGTACTGAGCTCTGCTGCTGGGCGCTAATCACTGCTGATATGACAGCTGT 1656
Db 1594 AGTGGGCTGTACTGAGGCTCTGCTGAGCATCATATACCCCGCTGATATGATGATGAT 1653
QY 1657 GTTCTTCTAGGCTCTTCTGCTTCAACACCTGTGTGAGACACTACTGATATACC 1716
Db 1654 TGTGATCTA---TTCTGATGCGCATACCAACCTGATACATACAAATATGCTATCC 1710
QY 1717 GACTGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1776
Db 1711 AAATGTGCTTACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1770
QY 1777 TGGATGATTTTCTTGTACAAATACGATACCGGACCTTCAAGCAGATGATGATGATG 1836
Db 1771 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1830
QY 1837 AGCCTTCCACTCCAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1869
Db 1831 AGCCTTACCGCGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1863

```

```

RESULT 10
AF369383
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Aedes aegypti.
Aedes aegypti.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.
1 (bases 1 to 1996)
Dasher, M.K., Kohn, A.B., Harvey, W.R. and Stevens, B.R.
AEMAR, A Novel Amino Acid Transporter from Larval Aedes aegypti
Midgut
Unpublished
2 (bases 1 to 1996)
Dasher, M.K., Kohn, A.B. and Harvey, W.R.
Direct Submission
Submitted (09-APR-2001) Physiology, Whitney Laboratory, University
of Florida, 9505 Ocean Shore Blvd., St. Augustine, FL 32080, USA
Location/Qualifiers
1..1996
/organism="Aedes aegypti"
/db_xref="taxon:7159"
1..1848
/note="NAAT"
/codon_start=1
/product="amino acid transporter"
/protein_id="AAM13400.1"
/db_xref="GI:20152861"
/translation="MATVSYSSMHPSPRTOKIALSGRTSNEFLCHVFOCPWMSKRCGR
EPPTAENGGAPLIPYIVLFFVIGRPLYLEMALGGTSSSVKIEWEPLFEGIGI
GOLVETSVYSYSLALTHIFAFASDELPRATCKDNADNCVDSILVNMEDS
NSTGGQVSSQIYFLDLVLEKESIDGIGAPRMLTLILLANVYIPLIVYCVS
SGKAVFLAIPYVYLIIILRALTEADVGVEFTKPMGELNFKVMSATTOLE
FSLVGMSSIIIMFSYNNHNIYRDAMIYVTDFTSLGEMTIFSILGLAHNIDF
EDISVYSGTGLAFISYDIAKEDIVPOLSEVLFEMFLVGVSAVALHGAITIA
FMDAEPKRYWHLALISTIEGEGIVYIHLVANGYSIMSTTEAFSDLRALITEMVA
IFWYIGIDNKNNDIEPMVORVGLYVAMAGSNATPDCHREHFLIVEYKPTYSGOO
YPEALIGIGITVFIEGIIQVILIMAWYITGTSRESVWGRKSKAKPSOSMGPCTEHT
RKAWKYEENKSRDDIILYKNNHSYKRLCVLEGMIDETIQDSTRLL"
BASE COUNT
539 a 388 c 458 g 610 t
ORIGIN

```

```

Query Match 12.8%; Score 366.8; DB 3; Length 1996;
Best Local Similarity 54.5%; Pred. No. 2.4e-69;
Matches 936; Conservative 1; Mismatches 753; Indels 28; Gaps 9;
QY 239 GAACGCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 298

```

Db 45 GCAGCCCAAAATGCGCTGAGTGTGCAACAGCTGATTTCTTATGTCATGATTTTCA 104  
 Oy 299 ACATCCGTGGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 358  
 Db 105 ATGTCCGTGGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 162  
 Oy 359 GGTGCTTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 418  
 Db 163 GGAGCGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 222  
 Oy 419 TTAGAGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 478  
 Db 223 CTGGAAATGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 282  
 Oy 479 CCGGCGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 538  
 Db 283 CCAGCTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 342  
 Oy 539 TACGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 598  
 Db 343 TACGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 402  
 Oy 599 CCATGGGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 655  
 Db 403 CCGTGGGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 462  
 Oy 656 GCTGATGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 706  
 Db 463 AATGAATGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 522  
 Oy 707 TTTTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 766  
 Db 533 TTCTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 582  
 Oy 767 TGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 826  
 Db 583 TGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 642  
 Oy 827 GGAGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 886  
 Db 643 GCGGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 702  
 Oy 887 ATCACTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 946  
 Db 703 ATCACTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 762  
 Oy 947 GTCACTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1006  
 Db 763 ATCACTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 822  
 Oy 1007 GTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1066  
 Db 823 TGTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 882  
 Oy 1067 TTTGAGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1126  
 Db 883 TTTGAGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 942  
 Oy 1127 TTTGAGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1186  
 Db 943 TTTGAGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1002  
 Oy 1187 GAGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1246  
 Db 1003 GAGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1062  
 Oy 1247 GCGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1303  
 Db 1063 GCGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1122  
 Oy 1304 TCGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1363

Db 1123 TTTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1182  
 Oy 1364 GAGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1423  
 Db 1183 GATGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1241  
 Oy 1424 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1483  
 Db 1422 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1301  
 Oy 1484 TACGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1543  
 Db 1302 TACGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1359  
 Oy 1544 ATTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1603  
 Db 1360 ATTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1419  
 Oy 1604 GCTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1662  
 Db 1420 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1479  
 Oy 1663 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1722  
 Db 1480 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1535  
 Oy 1723 TGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1782  
 Db 1536 AGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1595  
 Oy 1783 TGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1839  
 Db 1596 TGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1655  
 Oy 1840 CTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1899  
 Db 1656 GGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1715  
 Oy 1900 GTGTGTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1937  
 Db 1716 ATCACTGTGTGTGGGTACGTCGTGGCGTTCCTTCATCGCGTACGAGATGGAGA 1753

RESULT 11  
 AC015177  
 LOCUS  
 DEFINITION  
 AC015177  
 Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 ACCESSION  
 AC015177  
 VERSION  
 HTG: 6436158  
 KEYWORDS  
 HTG: HTGS PHASE2.  
 SOURCE  
 Drosophila melanogaster.  
 ORGANISM  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydriidae; Drosophilidae; Drosophila.  
 REFERENCE  
 1 (bases 1 to 66809)  
 Adams, M. and Venter, J.C.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT  
 This sequence was identified as CDM:10211923 by the submitter.  
 For further information on this sequence e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

FEATURES  
 source  
 1..66809  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT  
 19034 a 14604 c 14832 g 18339 t  
 ORIGIN





Best Local Similarity 56.2%; Pred. No. 3.2e-49;  
Matches 595; Conservative 0; Mismatches 409; Indels 55; Gaps 2;

```

QY 720 TTCTCCAAAGGATGATGATGAGAGGTCTCGGTGCCCCCATCTGACTTGGTGT 779
    |||||
Db 7801 TTCTGCGGAGAAAGCGAGATGATGATGATGATGATGATGATGATGATGATG 7860
    |||||
QY 780 TGTGCTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 839
    |||||
Db 7861 TGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7920
    |||||
QY 840 CCGGCAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899
    |||||
Db 7921 CCGGCAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7980
    |||||
QY 900 TCACACATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 959
    |||||
Db 7981 TGGCGGACATCACTTCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 8040
    |||||
QY 960 GGGGCAAACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019
    |||||
Db 8041 GGCACAAAGCTCTGGAACCGAGGTCTGTATCGCGCGCGCGCGCGCGCGCGCGCA 8100
    |||||
QY 1020 TGACAGTGTGACCGGACCGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1079
    |||||
Db 8101 TGGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8160
    |||||
QY 1080 TCTA-----CAGG 1087
    |||||
Db 8161 TTATACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8220
    |||||
QY 1088 GATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1147
    |||||
Db 8221 GATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8280
    |||||
QY 1148 TTGGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1207
    |||||
Db 8281 TTGGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8340
    |||||
QY 1208 GCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1264
    |||||
Db 8341 AATGGAGGTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8400
    |||||
QY 1265 CAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1324
    |||||
Db 8401 CTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8460
    |||||
QY 1325 TCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1384
    |||||
Db 8461 AATGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8520
    |||||
QY 1385 ACCGTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444
    |||||
Db 8521 AACTGACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8580
    |||||
QY 1445 ACACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1504
    |||||
Db 8581 ACACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8640
    |||||
QY 1505 CTTTTCGCGCATTTTCTGAACTGCGAGGCGGTGTGTATGATGATGATGATGAT 1564
    |||||
Db 8641 CTGCTGTGCGCATTTCTGAGCTGTGATGATGATGATGATGATGATGATGATGAT 8700
    |||||
QY 1565 TGGCTAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1624
    |||||
Db 8701 TGGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8760
    |||||
QY 1625 GGGCTATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1684
    |||||
Db 8761 GCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8820
    |||||
QY 1685 AACCTGTGTGCGAGCAATCTGATATGCCGATGCT 1723
    |||||

```

Db 8821 CCCCTCAAGTACAAGATTAACCTACCTACCAATCGGCTGT 8859

RESULT 13  
AC005894  
LOCUS

DEFINITION  
AC005894 196337 bp DNA linear INV 31-OCT-1998  
Drosophila melanogaster, chromosome 2R, region 46A1-46B2, P1 clones  
DS00050, DS00191, and DS05369, complete sequence.

AC005894 AC005435 AC005431 AC005441  
AC005894.1 GI:3818342

KEYWORDS  
HTG.

SOURCE  
ORGANISM

Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazey, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummel, S.R., Karia, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomoth, M.A., Mazda, P.,  
Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacle, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Poon, L., Poon, E., Sequiera, A., Sethi, H., Sutr, E.,  
Svirskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R.,  
Zieran, L.L. and Rubin, G.M.  
Sequencing of Drosophila chromosome 2R, region 46A1-46B2  
Unpublished (1998)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

2 (bases 1 to 196337)  
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazey, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummel, S.R., Karia, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomoth, M.A., Mazda, P.,  
Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacle, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Poon, L., Poon, E., Sequiera, A., Sethi, H., Sutr, E.,  
Svirskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R.,  
Zieran, L.L. and Rubin, G.M.  
Direct Submission  
Submitted (31-OCT-1998) Berkeley Drosophila Genome Project, MS  
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US

COMMENT  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).  
P1 library locations: 1-50, 2-95, 56-89.

FEATURES  
source

1. 196337  
/organism="Drosophila melanogaster"  
/strain="Y2: cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2R"  
/map="46A1-46B2"  
/clone="P1s DS00050 (D349), DS00191 (D345), and DS05369  
(D355)"  
/clone\_lib="P1 library, partial Sau3A in pAD10SACB11"  
/note="These P1s, DS00050 (D349) DS00191 (D345) DS05369  
(D355) were completed as a project. DS00050 extends from  
P1 end at bp 1 to P1 end at bp 79763. DS00191 extends  
from P1 end at bp 68,333 to P1 end at bp 154,851 DS05369  
extends from P1 end at bp 109,192 to P1 end at bp  
196,337."

BASE COUNT  
ORIGIN

56314 a 42176 c 42152 g 55695 t  
Query Match 9.6%; Score 274.6; DB 3; Length 196337;  
Best Local Similarity 56.2%; Pred. No. 3.2e-49;  
Matches 595; Conservative 0; Mismatches 409; Indels 55; Gaps 2;

QY	720	TTCTCCAAACAAACGATGGAATTAAGAGAGCTCGTGGCCCCATCTGGTACTTGGTGT	779
Db	187505	TTCTGGCGCGAAGAGCGAGCATGATGATGATGCAATATCCGATCTGAGGCTGGCGC	187564
QY	780	TGTGTCTATTTCATCGCATGGCTCATGGTGTCTCGAGTCGTGGCCCCGAGAGTCAAGGTT	839
Db	187565	TGGCACTGGCTGTGGCTGGATGTCATGCGGGGAAATTATGTTCAAGGGGTCAAGAGTT	187624
QY	840	CGGGCAACCGGGCTACTTCTCGGCTCTTCCCATACGTTGTCATGATCATTATTC	899
Db	187625	CGGGCAAGGCTCTCTCACTCTCTCGCCCTCTTCCCTACAGGTGGATGCTGTCTCCTGG	187684
QY	900	TCACCAATTCATCTCGCCGGTGTACTAGAGGCAATCCGTTCTTCGTACAGCCCAAT	959
Db	187685	TGGGGCACTCACTTGGCGGGGCGCTTGCAGCGGCTGTGTACTTCTCGGCCCGCAAT	187744
QY	960	GGGCGAAATCTCTTGAGCTGGGTGTATGGTACTCAGCAAGTCAAGCAAGTTCCTCTC	1019
Db	187745	GGCAACAAGTCTCTGGAACCGCAGGTCGTGATACCCGCCCTCACCGAGGTCTTCTCA	187804
QY	1020	TCACAGTGTGCAACCGGACCGATCATGTTCCTTCCTACAAAGGTTTCAGCATATA	1079
Db	187805	TGGCCATCTGCTTCCGGGAACATCATATGTACGCTGTACAAACGCTTCGGCACAA	187864
QY	1080	TCCTA-----CAGG	1087
Db	187865	TTTACAGGTGAGTGGAGATATCCGGTTTGGAGTTCCTTACGTGATTCCTGCTCCACAG	187924
QY	1088	GATGCTGGATTTGTTACGACTTTTGACACCTTTACAGTTTCTTGTCTGGGTGCAGATC	1147
Db	187925	GATGCCATATCTGTACACAGCTGTGACACTTTCACCTCCCTACTGTCGGTGTGATATC	187984
QY	1148	TTGCGTATCTTGGTAACCTCGCTACGAACTCACTAGAGGTGGAGATGGTGGGT	1207
Db	187985	TTGCGGATTTCTGGGCAACCTGGCGTACGGAACCAACACCGACATCGCTAGTGGTTC	188044
QY	1208	GCTGGCGGACCAAGTCTTGTCTTCAATTATACCTGATGCCATTGCCAA---AACATTTC	1264
Db	188045	AATGGAGGTCGGGCTCGCTTTCATATCTTACCGGAGCCATTGCCAAAGTTTAAGTGG	188104
QY	1265	CAACCTCAGTATTTCTGGTGTCTTCCAGATGATGCTGGTGGTGGTTCACGCTCA	1324
Db	188105	CTGGCGCAGCTATTTCTCGGTGTCTTCTTCTCCTATGCTCTTGTCTGGGCATTCGGCAGC	188164
QY	1325	TCGCTGGCTCTGCTATGCACCTTTCACACATTGGGATGAGAGCGCTCCACGTTACCC	1384
Db	188165	AATGGGCGCATGGCCTCTCGATGTCACCCGATCAAGGATCAATTCGGAACCTGGAAG	188224
QY	1385	ACCGCTCATATGTGCGCATGACCTGTTCTTCCGGTTCTTCTGTGGACTTGTATTACGC	1444
Db	188225	AATCTGACTGTGGTGGTGGTATGACCATTTGAGGCTACTTCTCGGCTCTCTGTACATTC	188284
QY	1445	ACACCGGTTGACATATATTTCTTGAGCTTGTAGATCACTACGCGTGAACATTCCTGTG	1504
Db	188285	ACACCGGTTGCGAGTTCTCTGCTCAACCTGTCGACACTTGTGGGCTCACGTTCTGTGCGC	188344
QY	1505	CTTTTTCGCCCATTTCTGAACTGCGAAGCGGTCTTGAGTTATGATGGAGATCTG	1564
Db	188345	CTGTGTCTGCGCATCTTCGACTGTGTGACCAATTCGCTGATCTAGGTGTAAGGACTTC	188404
QY	1565	TGCTTACATATGATGATCTGTTGGGTAAAAAGACTGTGCTTACTGGCGTCTCGCTGG	1624
Db	188405	TGCGGTGACGTGGAGTTCATATGATTTGGCATCAAGACTTGGCTGTACTATGCAATCTGGCG	188464
QY	1625	GGGCTAATCACTCTGCTATATATGACGACTGTGTTCTTCAACGCTCTTCTGCTCTAAC	1684
Db	188465	GCGGTGCTCACTCTCTGCTATGATCTCAACATTCATCTCAACCTGTTCTGTACAG	188524
QY	1685	AACCTGGGTTGGGCAACTAGATATACCGCACTGTT	1723
Db	188525	CCCTTCAATGTAAGGATTTACCTTACCAATTCGGGTGTT	188563

RESULT 14	AE003832	242172 bp	DNA	linear	INV 28-JUN-2002
LOCUS	AE003832				
DEFINITION	Drosophila melanogaster 2R	section 18 of 74 of the complete arm,			
ACCESSION	AE003832	AE013599	AE002787		
VERSION	AE003832.3	GI:21627589			
KEYWORDS	fruit fly.				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridiidae; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 242172)				
AUTHORS	Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,G.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,V.G., Champe,M., Pfeiffer,B.D., Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J., Andrews-Planck,Koch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Boltsakov,S., Borrova,D., Botchan,M.R., Bouck,J., Brockstein,P., Brothier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dunn,P., Durbin,L.E., Domres,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,L.E., Evansglist,C.C., Ferrara,C., Ferrera,S., Fleischman,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gottrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heilmann,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Mel,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpem,G.H., Ke,Z., Kienilton,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kul,P.D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mathe,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nussken,D.R., Paclob,J.M., Palazolo,M., Pittman,G.S., Pan,S., Pollard,J.D., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,X., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Wodgett,T., Wolley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,N., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.				
TITLE	The genome sequence of Drosophila melanogaster				
JOURNAL	20196006				
PMID	10731132				
REFERENCE	2 (bases 1 to 242172)				
AUTHORS	Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferritera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Paclob,J., Parasas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stepleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.				
TITLE	Sequencing of Drosophila melanogaster genome				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 242172)				



```

OY 900 TCACCAATCATCTCCGCGGCTACTGACGGATCTCTTCTTCTGTCAGGCTCAT 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191021 TCGGGGCACTCATCTTCCGGGCGCTTCGACGGGCTGCTACTTCCGGCGCAT 191080
OY 960 GGGCAAACTCTTGAAGTCTGGTATGTTACTAGCAAGTCAAGTCTTCTCTC 1019
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191081 GGCACAAAGCTCTGGAAACCGCAGGTCTGTACGCCGCTACCCAGGTTCTTCTAC 191140
OY 1020 TCACAGTGTGACCGGACCGATCATCATGTTCTCTCTTACAGGTTTACAGATATA 1079
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191141 TGGCCATCTCTTGGGAACATCATATGTAAGCTCTGTAACCCGTTCCGACACAA 191200
OY 1080 TCTA-----CGG 1087
    |||
Db 191201 TTTAAGGTGAGTGAAGATATCCGTTTGAATCTTCACTGATCTTCTGCTCTCAGG 191260
OY 1088 GATGCTGATTTTATAGACATTTGACACCTTTACAAATTTCTGTTGGTGCAGATC 1147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191261 GATGCAATATCTGACACCGCTGACACCTTCACTCCCTGCTCGGTGTATATC 191320
OY 1148 TTCGATCTTGTATACCTGCGATGACACCTCACTCAAGTGGAGATGTGCTCGT 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191321 TTCGGATTTCTGGCACTGCGATGACGAGAAACACACCGACATGCTAGTGTGTC 191380
OY 1208 GCTGGGCTACCACTTCTCTTCAATTTACCTGATGCTGATGCTCA--AACATTC 1264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191381 AATGAGGCTCGGCGCTGCTTCAATCTTACCCGACGATGCCAAGTTTAAAGTG 191440
OY 1265 CAACCTCACTATTTCTGCTGCTGCTTCTCTGATGATGCTGCTGCTGCTGCTCA 1324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191441 CTGCGCAGCTATTTCTGCTGCTGCTTCTCTCAATGCTTCTGCTGCTGCTGCTG 191500
OY 1325 TCCGCTGCTGCTATGCTATGCTTCAACATTTGGGAGTGAAGGCTTCCGACGTAC 1384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191501 AATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191560
OY 1385 ACCGCTCAATGTCAGCATGACCTTCTGCGGTTCTGCTGCTGCTGCTGCTGCTG 1444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191561 AACTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191620
OY 1445 ACACCGGCTGACATATATCTTGAAGTGTAGATCACTACGCTGGAACATTCCTGTG 1504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191621 ACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191680
OY 1505 CTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191681 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191740
OY 1565 TGCCTAGCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191741 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191800
OY 1625 GGCCTAATCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191801 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191860
OY 1685 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191861 CCCCTCAAGTACAGATTTACACTCACTCACTCACTCACTCACTCACTCACTCA 191899

```

```

RESULT 15
AC018319/c 18567 bp DNA linear HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
DEFINITION pieces.
ACCESSION AC018319
VERSION AC018319.1 GI:5552872
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

```

REFERENCE
AUTHORS 1. (bases 1 to 18567)
TITLES Adams, M. and Venter, J. C.
JOURNAL Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10214336 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
location/Qualifiers
1..18567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 5705 a 3897 c 3763 g 5202 t
ORIGIN
Query Match 9.4%; Score 270.2; DB 2; Length 18567;
Best Similarity 58.0%; Pred. No. 2.8e-48;
Matches 498; Conservative 0; Mismatches 358; Indels 3; Gaps 1;
OY 696 CTCACCTCTCTTTTGAACAGTCTCCCAACAAAGGATGGAATTGAAGAGCTCTCG 755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6023 GCGACCTCTCTCTGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5964
OY 756 GTGCCCCATCTGCTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5963 GATATCCAGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5904
OY 816 TCGTCGCCCGGAGAGTCAAGGTTCCGCAAGGCGCTACTCTCTGCGCTCTTCCAT 875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5903 TCATCATCTCGGGAGTGAAGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5844
OY 876 ACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 935
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5843 ACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5784
OY 936 TCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 995
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5783 TTATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5724
OY 996 CAGTGCAGCAAGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1055
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5723 CCGTACCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5664
OY 1056 CTTCACAGGCTTTCAGCATATATCTACAGAGGATGCTGCTGCTGCTGCTGCTGCTG 1115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5663 CGTACAAACGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5604
OY 1116 CTTTCAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5603 CATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5544
OY 1176 AACTCACTGAGAGTGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5543 AATGGGACCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5484
OY 1236 CATACCTGATGCCATGCCAAA--ACATTCACACCTGCTGCTGCTGCTGCTGCTGCT 1292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5483 CTTATCCGATGCCATGCCATGCCAAAAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5424
OY 1293 TCCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5423 TCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5364
OY 1353 CATGGGATGAGAGCGCTTCCACGCTGACCCACCGCTCACTGATGATGATGATGATGATG 1412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5363 CCGTGCCTCAAGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5304
OY 1413 CTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1472

```

Db 5303 TGATAGATTCTCTGGTGGGATTGATTACATCACTCCCGAGGTCAACATATATCATCAGC 5244  
QY 1473 TTGTAGATCACTACGGTGAACATTCCTTGCTTTCTGCGCCATTCTGAACTCGCAG 1532  
Db 5243 TTATGACTTTCACGGCGCTTGTCTCCCTGGTATCGGCCATCTTGAGCTCATTG 5184  
QY 1533 GCGTGTCTGGAATTATG 1551  
Db 5183 CCGTGGGTGATCTACGG 5165

Search completed: July 26, 2003, 02:02:05  
job time : 7286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 22:31:53 ; Search time 621 Seconds  
(without alignments)  
103771.522 Million cell updates/sec

Title: US-09-991-458-1

Perfect score: 2860

Sequence: 1 ggcacgaggttactgttg.....ataaaaaaaaaaaaaa 2860

Scoring table:

IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	455.6	15.9	1956	23	ABLO3173
2	395.4	13.8	2036	23	ABLI4421
3	393.8	13.8	2205	22	AAD03450
4	391.8	13.7	2473	22	AAD09681
5	391	13.7	2013	23	ABLO5887
6	383.4	13.4	1992	23	ABLO7751
7	383.4	13.4	2030	22	AAD09683
8	378.6	13.2	2062	23	ABL28723
9	378.6	13.2	2108	22	AAD09684

c	10	274.6	9.6	4540	23	ABLO3172	Drosophila melanog
c	11	270.2	9.4	4385	23	ABLO7750	Drosophila melanog
c	12	247.4	8.7	5270	23	ABLI4420	Drosophila melanog
c	13	231.4	8.1	4301	23	ABLO5886	Drosophila melanog
c	14	229.2	8.0	6148	23	ABLO28722	Drosophila melanog
c	15	222	7.8	2817	19	AAV01590	Rat glycine transp
c	16	211	7.4	1983	13	AAQ28118	Human norepinephr
c	17	209.6	7.3	2728	18	AAV58853	cDNA encoding high
c	18	209.6	7.3	2728	19	AAV28861	Rat proline transp
c	19	209.4	7.3	1854	22	AAH28082	Nucleotide sequenc
c	20	209.4	7.3	1854	22	AAH28086	Nucleotide sequenc
c	21	209.4	7.3	3404	13	AAQ26683	Rat dopamine trans
c	22	207.8	7.3	1854	22	AAH28083	DNA encoding human
c	23	207.8	7.3	1854	22	AAH28087	DNA encoding human
c	24	207.8	7.3	1985	14	AAQ41172	Dopamine transport
c	25	204.8	7.2	3919	14	AAQ53413	Sequence of the P
c	26	204.8	7.2	3946	24	ABLO6342	Kidney cancer rela
c	27	201.4	7.0	2640	13	AAQ26664	BDAT. Bos laurus.
c	28	197.8	6.9	2278	14	AAQ41055	Rat 5HT transpote
c	29	196.4	6.9	2394	21	AAQ61440	Human glycine tran
c	30	196.2	6.9	2278	16	AAQ94016	Human glycine tran
c	31	196.2	6.9	2415	16	AAQ94017	Rat 5HT transpote
c	32	194.8	6.8	2394	21	AAQ61439	Human glycine tran
c	33	194.8	6.8	2394	21	AAQ91842	Human glycine tran
c	34	194.8	6.8	2394	21	AAQ92001	Human glycine tran
c	35	194.8	6.8	2397	19	AAV22932	DNA encoding a hum
c	36	193.2	6.8	2394	21	AAQ61441	Human glycine tran
c	37	193.2	6.8	2397	19	AAV22915	Human GLYT-2 trans
c	38	193.2	6.8	2397	19	AAV22917	Human GLYT-2 trans
c	39	193.2	6.8	2397	19	AAV22920	Human GLYT-2 trans
c	40	193.2	6.8	2397	19	AAV22921	Human GLYT-2 trans
c	41	193.2	6.8	2397	19	AAV22922	Human GLYT-2 trans
c	42	193.2	6.8	2397	19	AAV22923	Human GLYT-2 trans
c	43	193.2	6.8	2397	19	AAV22924	Human GLYT-2 trans
c	44	193.2	6.8	2397	19	AAV22925	Human GLYT-2 trans
c	45	193.2	6.8	2397	19	AAV22926	Human GLYT-2 trans

#### ALIGNMENTS

RESULT 1	ABLO3173	standard; cDNA; 1956 BP.
ID	ABLO3173	
AC	ABLO3173;	
DT	26-MAR-2002	(first entry)
DE	Drosophila melanogaster expressed polynucleotide seq ID NO 4001.	
XX	Drosophila: developmental biology; cell signalling; insecticide;	
KW	pharmaceutical; gene; ss.	
KW	Drosophila melanogaster.	
OS	Drosophila melanogaster.	
XX	WO200171042-A2.	
PN	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US09231.	
XX	23-MAR-2000; 2000US-191637P.	
XX	11-JUL-2000; 2000US-0614150.	
XX	(PEKE ) PE CORP NY.	
PA	Venter JC, Adams M, Li PMD, Myers EW;	
PI	WPI: 2001-655660/75.	
XX	P-PSDB; ABB59070.	
DR	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT		









AC AAD03450:  
XX 13-JUN-2001 (first entry)  
XX Drosophila melanogaster bioamine transporter 1 (BTL) homologue cDNA.  
XX  
XX  
XX Fruitfly; transporter protein; bioamine transporter 1; BTL; screening;  
XX therapy; molecular transport disorder; biopesticide;  
XX Invertebrate model; ss.  
XX Drosophila melanogaster.  
XX  
XX Key Location/Qualifiers  
XX CDS 159..2048  
XX /\*tag= a  
XX /product= "Fruitfly bioamine transporter 1 homologue"  
XX  
XX MO200118178-A1.  
XX  
XX 15-MAR-2001.  
XX  
XX 08-SEP-2000; 2000WO-US24598.  
XX  
XX 09-SEP-1999; 99US-0153461.  
XX (GENO-) GENOPTERA LLC.  
XX  
XX Ebens AJ, Keegan KP, Winslow JW;  
XX WPI: 2001-235196/24.  
XX P-PSDB: AAE00297.  
XX  
XX Drosophila melanogaster Bioamine Transporter 1 (BTL) nucleic acid and  
XX protein, useful in screening assays to identify candidate compounds  
XX which are potential pesticide agents or therapeutics that interact with  
XX BTL proteins -  
XX  
XX Claim 2; Page 48-49; 53pp; English.  
XX  
XX The present sequence is a Drosophila melanogaster transporter protein  
XX homologue, bioamine transporter 1 (BTL). The BTL is used in screening  
XX assays to identify candidate compounds which are potential pesticide  
XX agents or therapeutics that interact with BTL proteins. It can also be  
XX used to genetically modify metazoan invertebrate animals resulting in  
XX BTL expression or mis-expression. It can also be used in methods for  
XX identifying new drug targets, therapeutic agents, diagnostics and  
XX prognostics useful in the treatment of disorders associated with  
XX molecular transport across the membrane. The BTL nucleic acid or its  
XX fragments, such as an antisense sequence or double stranded RNA, may be  
XX used as a biopesticide to inhibit BTL function. The invertebrate model  
XX organisms such as Drosophila melanogaster are useful for rapidly carrying  
XX out large-scale systematic genetic screens and therefore for analysing  
XX expression and mis-expression of BTL protein.  
XX  
XX Sequence 2205 BP: 478 A; 517 C; 570 G; 640 T; 0 other:  
XX  
XX Query Match 13.8%; Score 393.8; DB 22; Length 2205;  
XX Best Local Similarity 54.1%; Pred. No. 1,7e-76;  
XX Matches 872; Conservative 0; Mismatches 732; Indels 9; Gaps 3;

Db 516 TTCTCGAGCGGTGCTTGCATCAGAGCCTCGATATGACCTATATAGGGGTATTTGCC 575  
Qy 500 TACGTCACAGTCCGCGCTGCGGTACATCTCTCTTACTAGTGGTATCTGTGCTC 559  
Db 576 TATGTCAGGTGATTTCCACCGCTTGGCCACCACCTGATACCTCTGATTTAGGCTTG 635  
Qy 560 TGTCTATTTACTTACTATGACCTTCAAGGCGACCTTTCATGGGCTATTTGTGACGCT 619  
Db 636 ACCATGATATTTGTTGGCTTGTCTATGTAAGTCTGCTGACATATTTGCTCTG 695  
Qy 620 GAGTGGAGAACTGCTTACCTCGATCCAGAACACTTGTGATCACTGACATCAACATCAC 679  
Db 696 GATGGGGCAAGTGTGCGGAGGAGGAGCCAGAGCTGCCAATGATCATCATCTG 755  
Qy 680 AATGTACACAGAGTGTCTCACTCTTTTGTAGAACAGTCTCCAAACAGCGATGA 739  
Db 756 CAGGGGGTTTCCGCCCGAGCTGTTTTCACCAACAGTTCTGAGGAGGCCGAATCG 815  
Qy 740 ATTGAAG---GAGGTCCGGTGGCCCATCTGCTACTGTTGTTGTTGTTATTCATCGCA 796  
Db 816 CTGAGCATATATGGCCTGGGTACTCCAGTTGGGATCTGTTCTTGTCTCTGGCCACT 875  
Qy 797 TGGCTATGTTGTTGGAGTCTGCGCCGAGAGTCAAGAGTCCGCAAGCGGCTTAC 856  
Db 876 TGGGTATCATTCGACAGATTTTGTCAAGAGCATTCGCGAGTTGGAAAGCTCTAT 935  
Qy 857 TTCTCGCGCTCTCCCATACGTTGATGATCATCTTATTCATCAACCAATCATCTG 916  
Db 936 TTCTCGCGCTCTCCCATACGTTGATGATCATCTTATTCATCAACCAATCATCTG 995  
Qy 917 CCGGTCGCTACTACGAGCATCTGTTCTGTCAGCCCTCAATGGGCGAAACTCTTGA 976  
Db 996 CAGAGTCCCTGGAGGAATTTGTACTTCTTAAGCCCAATGTGCCAGCTGTCAAT 1055  
Qy 977 CTGCTGATATGTTACTACAGATGACGAGCAAGTCTTCTCTGTACAGTGTACCCGA 1036  
Db 1056 CCCACAGTGTGTAACGGGGCATACAGATATTTCTTCGCTGGCCATCTCTCGGA 1115  
Qy 1037 CCGATCATGATTTCTTCTTCAACGGTGTGACATATATTTCTAATATGCTTGG 1096  
Db 1116 AGCGTGTATGTAACGCTCTCTTCAATGATTTCAACAGATGTGCAAGATGTGATA 1175  
Qy 1097 ATTGTACGACTTGGACACCTTACAAAGTTCTGTGTTGGTGCAGATCTTGGTATC 1156  
Db 1176 ATATATCCACTATAGACCTGCTTACATTCATCTGGCTGTGATCATCTTGGAT 1235  
Qy 1157 CTGTGTAACCTCGCTACGACTCACTCACTGAGAGTGGAGATGTGCTGCTGGCGGT 1216  
Db 1236 CTGGGCAACCTGGCTTTGAGACCAACCAAGATATTTGCGAGGTTGCAAGGTGA 1295  
Qy 1217 ACCAGTTCCTTCAATTTATACCCGTGATGCCATTCGCAAA---ACATTCGAACCTAG 1273  
Db 1296 GCTGTCTTACGTTATCTTCTTATCCAGAGGCCATTCGCAATTTCAAGTATCTCCCA 1355  
Qy 1274 CTATTCGCTGCTGTTCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1333  
Db 1356 CTGTGCGCGCTTCTTCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415  
Qy 1334 CTGCTATGACCTTCAACACATGCGATGAGAGCGCTTCCAGCTGATACCCAGCTTAC 1393  
Db 1416 ATGGCTTGGCTGTGTAATGTGTAAGATGATTCACCAATTTGCCCCACTGGCTG 1475  
Qy 1394 ATGTACGAGTACCTGCTTCTGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1453  
Db 1476 CTGGCGGTAGTGTGATATGATTTCTCTGTGCGCTGTGCTGCTGCTGCTGCTGCTGCT 1535  
Qy 1454 GACCAATATATTTCTGAGCTTGAATGATGATGATGATGATGATGATGATGATGAT 1513  
Db 1536 GAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1595  
Qy 1514 GCAATTTCTGAACTCGAGCGGTGCTGATTTATGATTTGAGATCTGTGCTGATGAC 1573  
Db 1514 GCAATTTCTGAACTCGAGCGGTGCTGATTTATGATTTGAGATCTGTGCTGATGAC 1573

CC candidate compound especially a putative pesticidal or pharmaceutical  
CC agent that interacts with an invertebrate symporter cell surface  
CC receptor protein or its fragment. Insect or worm genetically modified  
CC to express protein of the invention is useful for studying invertebrate  
CC symporter cell surface receptor protein activity, by detecting the  
CC phenotype caused by the expression or mis-expression of the protein in  
CC the animal. Nucleic acids encoding the invertebrate receptor protein or  
CC their fragments are useful as biopesticides. SNF nucleic acids are  
CC useful for generating mutant phenotypes in an animal model or living  
CC cells that are used to study the regulation of genes encoding the  
CC proteins which are useful as pesticide or drug targets. The genetically  
CC modified organisms or cells are useful in screening assays to identify  
CC pesticides or therapeutics and thus are useful in the identification of  
CC new drug targets, therapeutic agents, diagnostics and prognostics  
CC useful in treatment of disorders associated with ion channels. The  
CC nucleic acid molecules are also useful as hybridisation probes.  
CC The present sequence is a cDNA encoding *Drosophila melanogaster* (dm)  
CC K<sup>+</sup> coupled amino acid transporter of the SNF family which is  
CC referred as dmKSNF.  
XX  
S0 Sequence 2473 BP; 625 A; 593 C; 627 G; 628 T; 0 other:

Matches 951; Conservative 0; Mismatches 767; Indels 69; Gaps 3

218 GATGACACGGACTTAGAGGCTGAACCGCCAGAACGTATGTTGTTCCACACACATTGAA 277

32 / GAAAAAACAGATGCCCGAGAGCCGACGGCGGAGCCACCAATTGGGCAATGGCCTGGAG 386

2/8 TTTCTTGGATGTCCCTGCATCCGTCGGTTTGGGTAACGTGTGGCGGTTCCCTTTC 337

20' AATCTTATTGCGAGTTCCCGTTC 446

338 ATGGCGTACCGAATGGAGGAGGCTTCTCCTGGTGCATACGTCATCGTCTTTACTT 397

44 / ACGGCTATGAGAAATGGAGGAGGCGCCTTCCTTATACCTACATCATAGTGTCTTCTG 506

27 536 GTCGCAAGCCTGCTGACIACITAGAGTGTCTCCGACCAATTCAGTTCAGAACTCT 457

...GAGGGCCAGGACG 560

[illegible]

**QUESTION**

[illegible][illegible]

Db  
687 GTGCTTCCAAATCGGAATTGCGGTGTCCTACTGTGCGGATGAGTGAACCAATTCCTC 745

QY 638 CCCTC----- 6A2

Db 747 AACTGAGGCCGAGGAGTATGTGACAATCTGCTTACGGGGTATCCCTTGGCCAATGAA 806

643 -----AGATCCAACTTGCTGCATCAGTCAACACATCACC AATGTTAC CAGCAGT 694

DJ 8V / JCGCCAGAAATCTCAGCGCATTTGTGGCCAACGATGAGACAGAGAACTCCAGAGCAGC 866

[illegible][illegible][illegible]

815 **ՀԱՅԱՍՏԱՆԻ ՀԱՆՐԱՊԵՏՈՒԹՅԱՆ ԱՆՏՀԱՆՈՒԹՅԱՆ ԱՆՏԱՐՈՒՄԻ ՎԵՐԱԿԱՆԱԿՈՒՄԻ ԿՈՄԻՏԵ**

[illegible]

```

QY 875 TACGTGTGATGATCACTTATTATTCACCAACATACCTGCCGGTCTACTGACGGC 934
   |||||
Db 1047 TACGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1106
QY 935 ATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 994
   |||||
Db 1107 ATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1166
QY 995 GCAGTCACGCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1054
   |||||
Db 1167 GCAGTCACGCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1226
QY 1055 TCTTACACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1114
   |||||
Db 1227 TCTTACACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1286
QY 1115 ACCCTTACAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1174
   |||||
Db 1287 ACCCTTACAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1346
QY 1175 GAACCTACAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1234
   |||||
Db 1347 GAACCTACAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1406
QY 1235 TCATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1291
   |||||
Db 1407 TCATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1466
QY 1292 TTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1351
   |||||
Db 1467 TTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1526
QY 1352 ACATGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1411
   |||||
Db 1527 ACATGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1586
QY 1412 TCTTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1471
   |||||
Db 1587 TCTTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1646
QY 1472 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1531
   |||||
Db 1647 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1706
QY 1532 GGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1591
   |||||
Db 1707 GGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1766
QY 1592 AAAAAGACTGCTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1651
   |||||
Db 1767 GCGCGAGTTCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1826
QY 1652 ACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1711
   |||||
Db 1827 ATCATATTCATATTCATATTCATATTCATATTCATATTCATATTCATATTCATAT 1883
QY 1712 TACCGAGCTGCTGCTTATGCTTCTGATGATGATGATGATGATGATGATGATGATG 1771
   |||||
Db 1884 TACCGAGCTGCTGCTTATGCTTCTGATGATGATGATGATGATGATGATGATGATG 1943
QY 1772 CCAATTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1831
   |||||
Db 1944 CCAATTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2003
QY 1832 AAGAAAGCTTCCACTCCAAACCTCCTACAGGCTGCTGCTGCTGCTGCTGCTGCTG 1891
   |||||
Db 2004 AAGAAAGCTTCCACTCCAAACCTCCTACAGGCTGCTGCTGCTGCTGCTGCTGCTG 2063
QY 1892 TGGATGAGTTCAGAGGCTGAGAGCGAAGCTCTTAGACAAAGATGAA 1938
   |||||
Db 2064 TGGATGAGTTCAGAGGCTGAGAGCGAAGCTCTTAGACAAAGATGAA 2110

```

```

RESULT 5
ABL05887
ID ABL05887 standard; cDNA; 2013 BP.
XX
AC ABL05887;
XX
DE 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 12143.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EM;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB61784.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12143; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2013 BP; 414 A; 511 C; 553 G; 535 T; 0 other;
XX
XX
Query Match 13.7%; Score 391; DB 23; Length 2013;
Best Local Similarity 53.2%; Pred. No. 6.6e-76;
Matches 949; Conservative 0; Mismatches 765; Indels 69; Gaps 3;
QY 222 ACACGAGCTTAGAGCTGAACCCGCGAAGACTATGATGATGATGATGATGATGATGATG 281
   |||||
Db 170 AACAGATGCGAGAACCCGCGAGGCGACCAATTTGGGCAATGAGCTGAGCTTC 229
QY 282 TGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 341
   |||||
Db 230 TGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 289
QY 342 CGTACGAGATGAGAGAGTGTCTCTCTGTCATGATGATGATGATGATGATGATGATG 401
   |||||
Db 290 CGTACGAGATGAGAGAGTGTCTCTCTGTCATGATGATGATGATGATGATGATGATG 349
QY 402 GCAAGCTGTGTACTACTTGAAGTGTCTCTCTGACATTCAGTTCAAGAACTGTGTA 461
   |||||
Db 350 GCAAGCTGTGTACTACTTGAAGTGTCTCTCTGACATTCAGTTCAAGAACTGTGTA 409
QY 462 AAGTTTGCTCATTTTACCGCGCATGAAGATGATGATGATGATGATGATGATGATG 521

```

D	b		410	AGATTGGCGGGTGGTGGCGGGGATTTGGGGGGTGGGCGATGAGCCAGGCGCTTGGCACC	469
O	y		522	GTTACATCCTGCTTACTACTACGTGGTGAATCTGTGTCTGTCTGTATTAATTAAGTATGA	581
D	b		470	TCTGATCATCTCGTACTATTCCTCGCTGTGGCCCTGACCTTACTATATCTTTGTGT	529
O	y		582	GCTTCAGAGCCACTTTCGATGGGGCTAATTTGTCACGCGAGTGGGAGAACTCCGTAACCT	641
D	b		530	CCTTCGAAATCGGAATGGCCGTGGTCTTACTGTGCGGATGAGTGGACCAATTCGTGAAC	589
O	y		642	C-----	642
D	b		530	CGAGGCCCGAGAGATGTGACAAATCTGCTTACGGGGGTATCCTTGGCGAATGAATGG	649
O	y		643	----AGATCCAAACACTTTCGTGATCAGTCAACATATACCATGTGTACCGAGTCTC	698
D	b		650	CCAGAAATCTCAGCGGCATTTGTGGCCACGATGAGAACAGAGAACTCCAGACACCTCG	709
O	y		699	AACTTACTTTTGGAAACAGTTTCCAAACAAGAGATGGAATTAAGAGAGTCCGGG	758
D	b		710	AACTTATTTCTGAAATGATGATGATCAAGAGAACTGTGACATCTCGGACCGCTTGGTG	769
O	y		759	CCCCATCTGTACTTGGTGTGTGTCTATTCATTCAGATGCTCATGCTGTCCGAGTCC	818
D	b		770	ATCCCGATGGAAAGTGAACCTTGCCTCCCTGTGCTGCGCTGGGTGTCAATCTTCTGAGTA	829
O	y		819	TCGCCGAGAGATCAAGAGTTCGCGCAAGCGGCTTACTTCCTCGCGCTCCCATACG	878
D	b		830	TCATCCGAGGGGTGAAGATTCGCGCAAGCGGCTTACTTCTGCGCTGTCCCTC	889
O	y		879	TTGTATATATCACTTATTCATCACCAATCATCTCCCGGTCTACTGACGACATCC	938
D	b		890	TGTGTCTTTCGTCTCTGCTGATCCCTGCGCTTACTTGGAGGAGACAGCATGTGCATTC	949
O	y		939	TGTTCTTCTGACGCTCAATGGGGGGAATCCTGACTGGGTGATAGTACTACGAG	998
D	b		950	TGTTCTTCTGAGGCCACAGTGGGTGACTCTTGAAACCCACGCTTTGGAAGAGGCG	1009
O	y		999	TCACGCAAGTGTCTCTCTGTGACAGTGTGCACCGGACGATCAATCATGTCTCTCT	1058
D	b		1010	TTGTGTCATGCTTCTTCTCTGCTGGCGCGGGGCTCGGACCATTAATCATGTTCCGCTGT	1065
O	y		1059	ACAACGGTTCAGACATATATATCTACAGGATGCTTGATGTGTACGACTTTGACACT	1118
D	b		1070	ACAATCGATTCGATCAGGAAATCTATAGGATGTCATATGTAAACACACATGCGACACG	1129
O	y		1119	TTTCAAGTTCCTTGTCTGGGTGACAGATCTTCGATCTTGGTATTCCTGGGTACGAG	1178
D	b		1130	TGACACGCTCTCTGGGAGGTAAACGATATTTGCATCTTTGGGCATTTGGCGGCATTC	1188
O	y		1179	TCAACTCAGAGGTGGAGATGTGTCGTGCTGCGCGGTACACAGTCTTCTTCATTTGAT	1238
D	b		1190	TGCAGATCGAAACATCAGAGATGTGGTGCAGAAAGTGCACGGGATTTGGCTTCAATCTGT	1249
O	y		1239	ACCGTGAATGCATTGGCCAA---AACATTCGAACCTCACCTATCTCGGTGCTCTCC	1295
D	b		1250	ATCCGGAATGCAATTCAGAGTTCCAGGCGGTGCGCACACTTTTCTCGGTTCTGTCTCT	1309
O	y		1296	TGATGATGTGGTGTGGGTATCGGGCTATCCGTGCTGTCTGTATCGACTTTCACACAT	1355
D	b		1310	TCATGCTGTTCTCTCTGGGCAATTCGATCAATTTGTGCGCTTCGACAGACACAAATTCACCA	1369
O	y		1356	TGGGATGGAGCGCTCCACAGTGAACCCACCGCTCTAATGTCAACGATGACGTGTCT	1415
D	b		1370	TCATTTGGGATCAGTTTCAAGGATGGAAGTCTGGAAGGTGGCACTGACCACTCTGTGT	1429
O	y		1416	GGGTTTCTCTTGGACTTTTACTGACACACCGGTGGGCAATATATTTTGAAGCTTG	1475
D	b		1430	GGGTTTCTTATGGGTTTGTGTACGTTACACCGGGGAGCAAGTGAATCTCACTTTGG	1489
O	y		1476	TAGATCACTACGTTGGAAATCTCTGTGTCTTCTGTGCGCATTTTGTGAATCGAGCG	1535

Db 1490 TGGACTTCTAAGCGGGAACCTATGCTGGCTTCATCTTGGCCATTTTGGACCTGGCTGGCA 1549

QY 1536 TGTTCGTGATTATGATTTGGAGATCTGTGCTAGACATGATGATTCATGTTGGGTAAAA 1595

Db 1550 TCGTGTGGGTTATATGTCCTGCMAAACCTTTTCGATGACATGAGTTTATATGTGCATAGCC 1609

QY 1596 AGACTGATGCTTACTAGGCGCTCTGCTGGGGCGTAATCACTGCTGCTTAATGACGACTG 1655

Db 1610 GAGTTTCACTACTAGCGGAGTATGCTGTGCTCTTCTACGCGCGGTATGATGATCA 1665

QY 1656 TGTTCCTTCACGCTCTCTCTCGCCTTACAACACTGTGTGCGAGACACTACGTATACC 1715

Db 1670 TATTCATTAATCTCGATGGTGTACACATCGAACCGGATCAATATATAGGAACCTGTAC---TTCC 1726

QY 1716 CGACTGCTGTTATGTTTCTGTGATCACTGATGTTATATTTTGGCATGACGTTTGCCAA 1775

Db 1727 CCGAAGCCCGCAACATTTCCGGCTGGGCTGTTCGCCATGGGTCCTCCACAGTTTCCAC 1786

QY 1776 TTGGAATTTGGATTTCCTTTGTACAAATACGCTACCGGACCTTACGCGACGATCAAGA 1835

Db 1787 TGTGGGGGCTTTGGTACATCTCGCGCCATCCACAAGGACGCTACTGTGAAAGTCTGTGAAG 1846

QY 1836 AAGGCTTCACATCCAAACCCATGAGGGGTCCGCCCTCGCCGAGAGAGGTGAGAAATGGA 1895

Db 1847 CCTGCTGAACCCCAAGCGATCGCTGGGGTCTCTGCCAATCCGGAATTTCCGGCGCAATGGG 1906

QY 1896 TGCATTTCAAGCTGGAAGCGAAACCTCTTACACAAGATATAA 1938

Db 1907 TGAATCTTCAAGATCAGAGGCGAGCCCAAGAACGCGACGACAGA 1949

RESULT 6  
ABL07751  
ID ABL07751 standard; cDNA; 1992 BP

AC	ABL07751;
XX	
DT	26-MAR-2002 (first entry)
YY	

Drosophila melanogaster expressed polynucleotide SEQ ID NO 17735.

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.

05 *Drosophila melanogaster*.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.  
XY

PR 23-MAR-2000; 2000US-191637P  
DP 11-TUT-2000; 2000US-0614150

XX  
XX  
DA (DEFE) DE CORR NY

XX  
XX  
Venter JT. Adams M. Li BMD. Myers EW

XX WPT: 2001-656860/75  
DR

DR P-PSDB; ABB03048.  
XX

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

XX  
XX  
**INTERACTIONS**  
XX

Citation 1, seq id no 1730; zipp + sequence listing; English

is the invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention







KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN MO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001MO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 37642; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA  
 CC sequences (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 2062 BP; 497 A; 496 C; 511 G; 558 T; 0 other;  
 Query Match 13.28; Score 378.6; DB 23; Length 2062;  
 Best Local Similarity 53.28; Pred. No. 3.5e-73;  
 Matches 976; Conservative 0; Mismatches 799; Indels 58; Gaps 6;

Db 443 GGCCTTCAGAGTGTTCATGCTCAGCACTACTATGCGCCCTGGTAGCATTAATTTGGCCG 502  
 QY 568 TTACTTATGCTATGAGCTTCCAGGCCACTCTTCCATGGGCTATTGTTCAGCTAGTGGGA 627  
 Db 503 ATATTTCATTGAATCATTTGCAACCCATTGGCCCTGGTCCACATGCGTGGGAATGGGG 562  
 QY 628 GA---ACTGGCTACCCCTCAGATC----- 647  
 Db 563 TATCCACTGATTAATCACTACACCCCGATGGAGTAATTTGCTCACTGGAAAGCAATGA 622  
 QY 648 -CAACACTTCTGATCATGCAACACATACCA-----TGTACACAGCATGCTCA 639  
 Db 623 TCAGCGCCCGCAAAATTAACAAATGAATACAAAAGATGGGGTCACTACAGTTCCGA 682  
 QY 700 ACTGTCTTTTGAAGCACTTCTCCACAAAGAGATGAATGAAGAGAGTCCGCTGC 759  
 Db 683 ATGTAATTTGTTAAGAAAGTGTGCGAGAAACCAATATATGAGAGAAAGAAATCGGCT 742  
 QY 760 CCCCATCTGTAATGTTGTTGTATTCATGCAATGCAATGATGATGTTTCGAGTGT 819  
 Db 743 ACCAACTGGAGTGTGCTATTTGAGCTTTTATCGCTGCTGTGTGTTTTCATTAAT 802  
 QY 820 CGCCCGAGAGTCAAGAGTTCCGGCAAGCGGCTTACTTCTCGCGCTTCCCAATAGT 879  
 Db 803 TCGTCGGGAGTGAAGAGTTCCGGAAAGGATCTACTTCTCGGCTTTCCGTCGT 862  
 QY 880 TGTGATGATACATTATATCATCAACATATCTCCGCGTCTACTACGCGATCCT 939  
 Db 863 CATCATGAGTGTGTTCTGTGCGAGGTGTAATCTGCGGATCATTAACAGGTATAATA 922  
 QY 940 GTTCTTCGTACGCTTCATGAGGCGCAACCTTGAAGCTGGTATGATGATGATGATGAT 999  
 Db 923 CTATCTTATTAACCGCGAATGGGAAAGATTTGGACCGCAAGTTTGGTACCGGCTGT 982  
 QY 1000 CAGCGAAGTCTTCTCTCTGACAGTGTGACCGGACGATCATATGTTCTCTT 1059  
 Db 983 AACTCAGTGTCTTACTCTTCTGTCCTGTTTCTTGGGCAATCATCATCTGATCTCTCT 1042  
 QY 1060 CAACGGTTTCAAGATTAATATATACAGAGATGCTGGATGTTACGCTTGGACACTT 1119  
 Db 1043 CAACAACTGGGACATATATACAGAGATGCAAGCAATTTGAACGGGTGTGGACACTAT 1102  
 QY 1120 TACAACTTCTTCTGCTGTCAGATCTTCCGATCTTGGTAACTCGCATGCAACT 1179  
 Db 1103 GACCTCTCTCTAGTGTGATTAACGATATTCGATATCTGGCCATCTGGCCACGAAT 1162  
 QY 1180 CAATCAGAGTGGGAGATGTGTCGTCGTCGCGGTACCAAGTCTTCTTCAATTCATA 1239  
 Db 1163 CGAACCAGATGACATGCGCTCACTTGTGAAGGCGGTGCGGATGCGCTTCAATCTATA 1222  
 QY 1240 CCTGATGCTTGGCA---AACATTTCAACCTCAAGCTATTTCTCGGTGCTTCTTCT 1296  
 Db 1223 TCCGATGCTATTTGCCAAGTTCAGATATCTGCCCAAAATTTCTGATGCTGTCTTCT 1282  
 QY 1297 GATGATGTGCTGCTGGGATATGCGCTATCCGCTCTGCTATGCACTTCAACATTT 1356  
 Db 1283 CATGCTCTCTGCTGGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1342  
 QY 1357 GCGGATGAGCGCTTCCACAGTATACCAACCGCTCTATATGATGATGATGATGATGAT 1416  
 Db 1343 CATTCGGGATGCTTCCCAACTTTGGGCGAGTGGCGAGTGTCTCTATTCCTCTAAT 1402  
 QY 1417 CGGTTCTGCTGCTGCTTCTTACTGACACCGGGGAGCAATATATTTCTAGGCTGT 1476  
 Db 1403 CAGTTCTTATTTGAT 1462  
 QY 1477 AGATCACTACGATGGAACATTTCTGCTTCTTCTGCGCATTTTGAATCTGACAGCGGT 1536  
 Db 1463 GGAATTTTTCGGGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1522  
 QY 1537 GTTCTGATTTATGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596  
 Db 1523 TGGATGATTTATGACACAGATGCTGTCGCAAGGACATTTGATGATGATGATGATGATGAT 1582



OY	1597	GACGAGCTCTAC	TGCGCCTCTCTGCTGGGCGTA	TCACTCTGCTATTAATGACAGCTGT	1656
Db	1583	AGTGGGCGCTGAC	TGGAGAGCTCTGTCGAGCATATTA	ACCCCGGTGATCATCATGTTAT	1642
OY	1657	GTTCTTCTACACCT	TTTTCGCGCTTAACAACTCGTGTTGGGAGACA	CTACGATATACC	1716
Db	1643	TCTGATCTGA---	TTTCTATGCGACCTACCAACCTCTACATACAA	CAATATGCTCTATCC	1699
OY	1717	GACGCTGCGTTAT	TTTCTGTGAGTACTTGATGTTATTTTGGGCAT	GAAGTTTTCGCCAAT	1776
Db	1700	AAATTTGGCTCTT	ACCAATTTGGTTGGCTAATTAACGGCCTTGCGTAT	CTTCAGCTGCCAAT	1759
OY	1777	TGGAATTTGATTT	CTTTGTGTAACAATACCGTAACCGTAACCTTACGCGAGACGATCA	AGAA	1836
Db	1760	TTGATGATGCTGG	CCATTTGTTGCAAGTCCAGGTCAACACTTTTGGGTGC	AAAATTCGCTGG	1819
OY	1837	AGCCTTCACATCC	CAAAACCTTCATGGGTCCCGG	1869	
Db	1820	AGCCTTTACGCCG	AAAAAAGATTTGGGGACCCAG	1852	

Novel invertebrate symporter cell surface receptor proteins and nucleic acid encoding the protein useful as pesticide or drug target and to identify compounds that have utility as therapeutics or pesticides -

receptor/protein or its fragment. Insect or worm genetically modified CC  
to express protein of the invention is useful for studying invertebrate CC  
symptoter cell surface receptor protein activity, by detecting the CC  
phenotype caused by the expression or mis-expression of the protein in CC  
the animal. Nucleic acids encoding the invertebrate receptor protein or CC  
their fragments are useful as biopesticides. SNF nucleic acids are CC  
useful for generating mutant phenotypes in an animal model or living CC  
cells that are used to study the regulation of genes encoding the CC  
proteins which are useful as pesticide or drug targets. The genetically CC  
modified organisms or cells are useful in screening assays to identify CC  
pesticides or therapeutics and thus are useful in the identification of CC  
new drug targets, therapeutic agents, diagnostics and prognostics CC  
useful in treatment of disorders associated with ion channels. The CC  
nucleic acid molecules are also useful as hybridisation probes. CC  
The present sequence is a cDNA encoding *Drosophila melanogaster* (dm) CC  
SNF homologue which is referred as dmSNF3. CC  
XX  
Sequence 2108 BP; 524 A; 501 C; 518 G; 565 T; 0 other;

CC The invention relates to invertebrate symporter cell surface receptors  
CC of the sodium/neurotransmitter family (SNF) and nucleic acid molecules  
CC encoding such receptors. The SNF protein is useful for detecting a  
CC candidate compound especially a putative pesticidal or pharmaceutical  
CC agent that interacts with an invertebrate symporter cell surface

Db	754	ACCAAACTGGGAGTTGGTCAATTGGACTTTTATTCGGCTGGTCTGGTGTTTTCAATTA	813
Qy	820	CGCCCGGAGTCAAGAGTTCCGGCAAAAGCCCTACTTCTCGCGCTCTTCCATACGT	879
Db	814	TCCGTCGGCGAGTAGAGTTCCGGGAAAGGCATCTTACTTCTCGCCCTTTCCGATAC	873
Qy	880	TGTCATATCATCTTATTTCAATCAACAAATCATCTCGCCCGGGTCACTACGAGCCATCC	939
Db	874	CATCATGGGTGTTCTTGTGGTGGAGCTGTAACTCTCGCGGATTCATTAAGCGTATTA	933
Qy	940	GTTCTTCGTCAAGCCTTCATGGCGAAATCCTTGAGCTTGGTATAGTACTACAGAGT	999
Db	934	CTACTTTATTAAGCCGCAATGGGAAAGATTTTGGATCCGAAAGTTTGGTACCGCGCTGT	993
Qy	1000	CACCGAAGTGTCTCTCTCTGACACGTGTGACCGAGCCATATATATGTTTCCTCTTA	1059
Db	994	AATCAGATGTTTCTACCTCTTCCGCTTTTCTTTCGGCAACATCATCTGTACCTCTGTT	1053
Qy	1060	CAACGGTTTCAGACATATATATCTACAGAGATGCTTGATTTGTATACACTTTGGACACT	1119
Db	1054	CACAAAGTTGGCGACATGTATACACAGAGATGACAAATTTGTAAAGGGTCTGGACATAT	1113
Qy	1120	TACAAGTTTCTTGTCTGGGTGACAGATCTTGGTATCTTGTGTAACTCGCGTACGAACT	1179
Db	1114	GACCTCTCTGCTACCTGATTCACAGATTTCCGATCTCTGGCCATCTGGCCACGAAAT	1173
Qy	1180	CAACTCAGAGTGGGAGATGTGTCGGGTGGCGGGTACACAGATTCCTTTCATTTTCATA	1239
Db	1174	CGGAACCGATACATCGCGCTCAGTGTGTGAAGGGCGGTCCGGANTGGCCTTCATATCATA	1233
Qy	1240	CCCTGATGCCATTTGCCAA--ACATTTCAACCTCAGCTATCTCGGTGCTGTCTTCT	1296
Db	1234	TCCCGATGCCATTTGCCAAATTCAGAAATCTGCCAAATCTCTCAGTGGTGTCTTCTCT	1293
Qy	1297	GATGATGCGAGTGGGGGTATCGGCTCATCGGTGCGCTGTGATACACTTTCAACACATT	1356
Db	1294	CATGCTCTGTGTGTGGGCAATAGATGGAATATAGCCATAGCCTCGTGTGGTACGCG	1353
Qy	1357	GGCAGTGGACGCGTTCCACGCTGATACCCACCGCTCTACATGTCAAGCATGACCTGTTTG	1416
Db	1354	CATTGCGGANTGGTTTCCCACTTTGGGCAATGGCAGTGGCTCGCTCATCGCTGTAGT	1413
Qy	1417	CGGTTTCCTGCTGGAGCTGTTTACGTGACACCGGGTGAACAATATTTCTGAGCTGT	1476
Db	1414	CAGTTTCTTTATTTGGATTATGTATATACACCGGGCGGTAGTATATGCTGACCTTGT	1473
Qy	1477	AGATCCTACAGCGTGGAACTTCTTGCTGTCTTCTTTCGGCCATTTCTGAACCTGCAGCGT	1536
Db	1474	GGACTTTTTCGGCGCTCAATGATTTGCTTGTACTGGGAATTCGCCAGCTGTACACCAT	1533
Qy	1537	GTTCTGGATTTATGGATTGGAAATCTGTGCTGACATAGATTCATGTTGGGTAAAAA	1596
Db	1534	TGGATGGATTTATGGCACAGATCGCTGTGAAAGCACTTATAGTTCATGCTGGGTCCGA	1593
Qy	1597	GACTGATCTTACTTGGCGCTCTCGCTGGGGCGGTATCACTCTCTATATGACAGACTGT	1656
Db	1594	AGTGGCGCTGACTGAGAGCTCTGCTGGAGCATATTACCCGGTATCATGACTGTATAT	1653
Qy	1657	GTTCTTACAGCTCTTCTCGCCTCTAACACACTGTGTGTGGAGAACATCTAGTATACC	1716
Db	1654	TCTGTATTA--TTTCTATGGACCTTCAACACTCTAATACAAACATATATGCTATACC	1710
Qy	1717	GACTGCTGTTATATGTTTGTGATCTTATGTATTTTGGGCAATGAGCTTTGGCAAT	1776
Db	1711	AAATTGTCTTACTAATTTGGTGTATTAAGCGCTTTCGTGTATCTTCAAGCTGCCAT	1770
Qy	1777	TGGAATGGATTTTCTTGTACAAATACCGTACCGGAACCTTCAAGCGAGATCAAGAA	1836
Db	1771	TTGGATGATCGTGGCCATTTGTGAGATTCAGAGTCAGACTTGGGTGCAAAAATCCCTGG	1830
Qy	1837	AGCTTTCACATCCAAACCTCATGGGGTCCCG	1869
Db	1831	AGCCTTTACGCGGAAAGAAATTTGGGGACG	1863

ID	AB031772/c	standard; cDNA; 4540 BP.
XX	AB031772;	
XX	26-MAR-2002 (first entry)	
XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 3998.	
XX	Drosophila; developmental biology; cell signalling; insecticide;	
XX	pharmaceutical; gene; ss.	
XX	Drosophila melanogaster.	
XX	WO200171042-A2.	
XX	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US09231.	
XX	23-MAR-2000; 2000US-191637P.	
XX	11-JUL-2000; 2000US-0614150.	
XX	(PEKE ) PE CORP NY.	
XX	Venter JC, Adams M, Li PWD, Myers EW;	
XX	WPI; 2001-656860/75.	
XX	P-PSDB; ABB59069.	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	
XX	genes from Drosophila and for elucidating cell signalling and cell-cell	
XX	interactions -	
XX	Claim 1; SEQ ID NO 3998; 21pp + Sequence Listing; English.	
XX	The invention relates to an isolated nucleic acid detection reagent	
XX	capable of detecting 1000 or more genes from Drosophila. The invention is	
XX	useful in developmental biology and in elucidating cell signalling and	
XX	cell-cell interactions in higher eukaryotes for the development of	
XX	insecticides, therapeutics and pharmaceutical drugs. The invention	
XX	discloses genomic DNA sequences (AB01840-AB016175) and the encoded proteins	
XX	(ABB57737-ABB72072).	
XX	The sequence data for this patent did not form part of the printed	
XX	specification, but was obtained in electronic format directly from WIPO	
XX	at ftp.wipo.int/pub/published_pft_sequences.	
XX	Sequence 4540 BP; 1201 A; 1103 C; 1110 G; 1126 T; 0 other;	
XX	Query Match 9.6%; Score 274.6; DB 23; Length 4540;	
XX	Best Local Similarity 56.2%; Pred. No. 2.7e-50;	
XX	Matches 595; Conservative 0; Mismatches 409; Indels 55; Gaps 2;	
QY	720 TTCCTCAACAAGCGATGGAATTGAGAAGGTCTCGGCCCCCATCTGCTACTTGGTGT	779
DB	2382 TTCTCCGGGAGAGCGACATCGATGATGCGATCGCATATCCAGCTGGAGCTGGCGC	2323
QY	780 TGTGCTATTATCGATCGGCTCATGATGATGTTGCGAGTGTGTCGCCGAGAGTCAAGATT	839
DB	2322 TGGCACTGGCTGTGGCTGGATCGTCATCGCGGGGATTAATGTCGAAGGTTGAAGATT	2263
QY	840 CCGGAAAGGGCTACTTCTCTCGCGCTCTTCCATACGTTGTATCATCTTATTATCA	899
DB	2262 CGGGCAAGGCTCTACTTCTCGCGCTCTTCCATACGTTGTATCATCTTATTATCA	2203
QY	900 TCACCAATATCTCTCGCGGCTGTACTGACGCGCATCTGCTTCTGACGCTCAAT	959
DB	2202 TCGGGCACTACTTGTGCGGGGCTTTCGACGCGCTGTACTTCTGCGGCGCAAT	2143

OY		960	GGGGAACACTCCTTGGAGCGTGCTATGTGCATCCAGACGACGCCAGTGTCTTCCTC	1019
Dd		2142	GGCACAAGCTCTTGGAAACCGCAGGCTGTGTAGCGCGCGTACCAGGCTTTCTTCAC	2083
OY		1020	TGACAGTGTGCACCGCACCCGATCATCATGTCTCTCTTACCAAGGTTTTCAGACATATA	1079
Dd		2082	TGGCCATCTGCTGGGMAACATCATCATGTACGGCTGTACAAACCGCTGGGCCAACA	2023
OY		1080	TCCTA-----CAGG	1087
Dd		2022	TTTTACAGGTGAGTGGAGATATCCGGTTTGAGATTCTTACCTGATTCCTGCTCCTCAGG	1963
OY		1088	GATACCTTGGATTGTATACGACTTTTGGACACCTTTTCAAGTTTCTGTGCGGTGACATATC	1147
Dd		1962	GATCCGCAATATCTGTGACACACGCTGTGACACCTTCCACTCCTACTGTCCGGTGTATATTC	1903
OY		1148	TTGCGTATCCTTGGTAACTTCGCGTACGAACTCACTCAGAGTGGAGATGTGTGTGT	1207
Dd		1902	TTTCGGGATTTCTGGGCAACCTGGCTGTACGAGAACAACACCGACATCGCTAGTGTGGTC	1843
OY		1208	GCTGGCGGTACCACTCTTGCTTTTATTTATACCCCTGATGCCATTGGCAA---AACATTTC	1264
Dd		1842	AATGGAGGTCCGGGCGCTGGCTTTTATATCTTACCCGGACGCGCATTTGGCAATTTTAATGTGG	1783
OY		1265	CAMCCTGAGCTATTCTCGGTCCTCTTCTTCCTGATGATGTGGGTGGGATGGGCTCA	1324
Dd		1782	CTGGCGGACGATTTCTCGGTGCTTTCTCTCTCAATGCTCTTCGCTCTGGGCGATGGCGACG	1723
OY		1325	TCCGTGGCTCTGTATGACTTTCAACACATTTGGCGATGACGCGCTTCCCACGTGTACCC	1384
Dd		1722	AATTTGGGCAAGGCGCTCCTGCACTTCCACCGCTGATCAAGGATCAAGTCTGGCACACCTGTAAAG	1663
OY		1385	ACCGTTCATCATGTCCAGGAGTACCTGTCTTTCGSGGTTCCGCTTGGACATGTGTACTGC	1444
Dd		1662	AACCTGACCTGTGGTGTGGTATATGCCATTBTBAGGCTACTTCTGGGCGCTTCTGTACATTC	1603
OY		1445	ACACCGGGTGACAAATATATTTCTTGAAGCTTTAGATACTACGCGGTGANAATTTCTTGTG	1504
Dd		1602	ACACCCGGGTGSCAGTTCCTGCTCAACCTGTGCACTACTTTGGCGTACAGTTGTGTGGCC	1543
OY		1505	CTTTTTCGCGCAATTTCTGAACTGCGACGCGTGTTCGTGATTTATGCATTGGAATCTG	1564
Dd		1542	CTGTGTCTGGCGCATCTTCGACCTGGTGAACAATTCCTCGGAATCTACGGTGTGTAACCGACTC	1483
OY		1565	TGCGTAGACATTTGAGTTTCATTTGGGTAAAAAGACTGGTGTCTACTGGCGCTCTCGGTGG	1624
Dd		1482	TGCCGTGACGCGGAGTTTCATGATTTGGCATCAAGACTTGCCTGTACTATCCGATCTGTGGG	1423
OY		1625	GGCGTAACTACCTCTGCTATATATAGACACTGTGTCTTCTACGCTCTTCTTCGCCCTTAAC	1684
Dd		1422	GCGGTGCACACTCTCTGCTATATGCTCACCACTTCATCTCAATCAACACTGGTGTCTGTACGAG	1363
OY		1685	AACCTGTGTTCGAGACAACTACGTATACCCGACTGCT	1723
Dd		1362	CCCCCTCAAGTACAGATTAACCTTACCAAATCCGGGTGT	1324

RESULT 11	
ABL07750/c	
ID	ABL07750 standard; cDNA; 4385 BP.
XX	
AC	ABL07750;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 17732
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	W0200171042-A2.

XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
RX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PsDB; ABB63647.  
DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PR interactions -  
XX  
XX Claim 1; SEQ ID NO 17732; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-ABL16175) and the encoded proteins  
CC sequences (ABB57737-ABB72022).  
CC (ABB57737-ABB72022).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 4385 BP; 1367 A; 986 C; 884 G; 1148 T; 0 other;

Query Match	9.4%	Score 270.2	DB 23	Length 4385
Best Local Similarity	58.0%	Pred. No. 2.5e-49.3		
Matches	498	Conservative	0	Mismatches 358; Indels 3; Gaps 1
QY	656	CTCAACTCTACTTTTGGAAACAGTTCTCCAAACAAGGATGGAAATTGAGAGAGTCTCG	755	
Db	2652	CGCAACTCTACTTGCACAGCATCGTGGTAATGAAACGAAATTCCTGGAGGAGGCATCG	259	
QY	756	GTGCCCCCATCTGGACTTGGTGTGTCTATTATATGCAATGCTCATAGTGTTGGAG	815	
Db	2392	GATAATCCAGTGGTAGTTGGCTTTAATGCTGGGGATTTCGGGTCAACCCGTAAACATCAA	253	
QY	816	TCGTGCCCCGAGACTCAAGAGTTCCGGCAAGCGGCTACTCTCGCGCTTCCCAT	875	
Db	2532	TCATCATTTCCGGGAGTGAAGATTCAGAAAGCGGCTATGTCTCGGCCCTTTTCCGT	247	
QY	876	ACGTTGTCATGATCATCTTTATTTCATCACACATATCATCTGCGCGGTCTACTGACGCA	935	
Db	2472	ACGTTGTAAATTTTATCTCGTCCGAGAGACATCATCAATACAGAGTCCCTACGAGGAG	241	
QY	936	TCCTGTTCTTGCTACGCTCAATGGGGAACCTCTGAGCTGCTGTATGGTCTCAG	995	
Db	2412	TTATGTACTTTCTGCTCCACAGTGGGAAACCTTTTGAGCGCGAGGTGGTATCAATG	235	
QY	996	CAGTCCACGAAGTTCTTCTCTCGACAGTGTGCACCGGACCGCATCAATGTTCTCT	1055	
Db	2352	CCGTCACCGAGGTGTCTTCTCCCGCGCTGTCTGGTGGTATCAATCATGTCTCT	229	
QY	1056	CTTACACAGGTTTCAGACATATATCTACAGGATGCTTGGATTCTTACGACTTTGGACA	1115	
Db	2292	CGTCAACACCGTTGGTCAATGCTCTTCGGGAGCCAAATATATATTAACACACCCCTGGCA	223	
QY	1116	CCTTTACAAAGTTCTGTCTGGGTGCACGATCTTGTGATCTTGGTAACTCGCGGTACG	1175	
Db	2232	CATTCACCTCACTGCTCTCGGGGTGATTAATCTTTGGATCTGGGAAACCTGCGTCAAG	2173	
QY	1176	AAGTCACTCAGAGGTGGAGATGTGTGGTGTCTGCGGATCAAGTCTGCTTTTCATTT	1235	

Db 2172 AATCGGGACCAAGACATTCGACGATGCTTAAGGAGCGTCCGTGATGGCTTTCATCT 2113  
Qy 1236 CATACCTGATGCAATGGCCAA--ACATTCACACTGCAATCTCGGTGCTCTCT 1292  
Db 2112 CCAATCCGAGTCCATGCAAGTCCAAATGTTTCTCCTCAAGTGTCTCTCTCTCTCT 2053  
Qy 1293 TCCGTATGATGCTGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1352  
Db 2052 TCCCATGCTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1993  
Qy 1353 CATGGCGATGACGCGCTTCCACAGTACCCACCGCTACATGTCAGCATGCTGCT 1412  
Db 1992 CCGTGTCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1933  
Qy 1413 CTTCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1472  
Db 1932 TGTATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1873  
Qy 1473 TTGTAGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1532  
Db 1872 TTATGAGCTTTCACGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1813  
Qy 1533 GCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551  
Db 1812 CCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1794

RESULT 12  
ABLI4420  
ID ABLI4420 standard; cDNA; 5270 bp.  
XX  
AC ABLI4420:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 37742.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmacological; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001MO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT P-PSDB; ABB70317.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 37742; 21bp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
CC sequences (ABLI1840-ABLI16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
SQ Sequence 5270 bp; 1379 A; 1127 C; 1174 G; 1590 T; 0 other;  
Query Match 8.7%; Score 247.4; DB 23; Length 5270;  
Best Local Similarity 55.5%; Pred. No. 2.6e-44;  
Matches 561; Conservative 0; Mismatches 391; Indels 59; Gaps 2;  
Qy 749 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808  
Db 2932 GGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2991  
Qy 809 TTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 868  
Db 2992 GGCACGATTTTGTCAAGGCGCATTCGCGATTCGCGAAACCTTCCTTCTGCGCTGCT 3051  
Qy 869 TTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928  
Db 3052 TTTCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3111  
Qy 929 GACGGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988  
Db 3112 CAGGGATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3171  
Qy 989 TACTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048  
Db 3172 TACGGCGCATTCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3231  
Qy 1049 TTCTCTCTTACAAAGCTTTCAGCATTAAT----- 1078  
Db 3232 TACGGCTCTCTCAATGATTTCAACAGAAATGTCACAAATGATGATGATGATGATGATGAT 3291  
Qy 1079 -----ATTCAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1112  
Db 3292 TTTCAGATCACTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3351  
Qy 1113 ACACCTTTACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1172  
Db 3352 ACTCGCTTACATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3411  
Qy 1173 ACNACTCAACTGAGAGTGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1232  
Db 3412 TTGAGACCAACACCAAGATATTTCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3471  
Qy 1233 TTTCATACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1289  
Db 3472 TTTCCTATCCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3531  
Qy 1290 TCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1349  
Db 3532 TCTTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3591  
Qy 1350 ACAATTTGCGATGAGAGCGCTTCCACAGTACCCACCGTCTCATATGTCAGCATGCT 1409  
Db 3592 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3651  
Qy 1410 GTTCTGCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1469  
Db 3652 CGATTAATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3711  
Qy 1470 AGCTTGTAGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1529  
Db 3712 ATCTCGGACCTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3771  
Qy 1530 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1589  
Db 3772 TGGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3831  
Qy 1590 GTTAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1649  
Db 3832 ATGTGAAGACCACTTCTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3891





Best Local Similarity 54.1%; Pred. No. 8e-39;  
Matches 503; Conservative 0; Mismatches 415; Indels 12; Gaps 2;

```
OY 746 GGAGTCTCGGTGCCCATCTGTACTGTGTGCTCTATTCATTCGATGCGTCATG 805
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1372 GAATATCTGTGTGATGATGAGTGGCCCTTGCGCTTTCTGGCCCTGGGTGATT 1431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 806 GTGTTCGAGTGTGTGCCCGAGAGTCCGAGAGTCCGCAAGCGCCTACTCTCCG 865
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1432 GTATATGATGCGTGGCAAAAGAAATTAAGACATCAGAAAAGTGTACTTCACAGCC 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 866 CTCTTCCCAATGCTGTGATGATCACTTATTCATCACACATATCTCCCGTGTCT 925
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1492 ACCCTCCCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 926 ACTGACGGCATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 985
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1552 GGAGCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 986 TGTACTCAGCAGTACGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1045
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1612 TGGAAAGATGAGGCACTCATATTTCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGT 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1046 ATGTTCCTCTTCAACGGTTTCAGACATATATCTACAGGATGCTTGGATTGTACG 1105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1672 ACTCTTCTTCTTCAACAAATTCATACAACTGATACAGGACAGCTTAATTGTAAAC 1731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1106 ACTTGGACACCTTACAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1732 TGCACCAACAGTGCCTAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1166 CTCGCTACGAACTCACTCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1792 ATGGCCAAACGAGCGCAAGTCAACATGTAGAAATGTG---GCTGACCAAGAGGCCAGCATTT 1848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1226 GCTTTCATTTTCATACCTGTATGCAATGTCACAAAACATTCACCTCAGCTATTCGCTG 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1849 GCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1286 CTGTCTTCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1909 ATCTTTTCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1346 TTCACACATTTGGGATGAGCGCTTCCACGTATCCACCGCTTACATGTACAGCATG 1405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1969 ATGTGACCTTCATCTCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1406 AC-----CTGTCTTGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2029 ACCCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1457 CAATATATCTTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2089 ATCTACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1517 ATTTCTGAACTCGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2149 ATATTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1577 GAGTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2209 GAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1637 CCTGTATATGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2269 CCGACCATTTTAACGTTTATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 26, 2003, 00:00:40  
Job time : 637 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 23:24:48 ; Search time 3897 Seconds

(without alignments)  
11885.833 million cell updates/sec

Title: US-09-991-458-1

Perfect score: 2860  
Sequence: 1 ggcacgaggtactgttg.....ataaaaaaaaaaaaaaaaaa 2860

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estlin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hlc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hlc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estom: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.8	7.0	682	12	BF490939
2	163.6	5.7	704	13	BM589455
3	151.6	5.3	777	17	CNS0110L
4	149.6	5.2	3118	11	BC026339
5	145.2	5.1	798	12	BG780471
6	141.8	5.0	670	12	BF486192

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BF490939	682 bp. mRNA	BF490939	1								
AT27575	5prime AT Drosophila melanogaster adult testes pOTB1	AT27575	1								
Drosophila melanogaster	CDNA clone AT27575 5 similar to CG4476:	BF490939	1								
BF490939	located on: 3L 67A3-67A3:: 04/09/2001, mRNA sequence.	BF490939	1								
BF490939	GI:13756639	BF490939	1								
EST		EST									
fruit fly		fruit fly									
Drosophila melanogaster		Drosophila melanogaster									
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;		Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;									
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;									
Ephydroidea; Drosophilidae; Drosophila.		Ephydroidea; Drosophilidae; Drosophila.									
1 (bases 1 to 682)		1 (bases 1 to 682)									
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman		Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman									
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan		, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan									
, D., Frise, E., George, R., Gonzalez, M., Guartin, H., Harris, N., Li, P.,		, D., Frise, E., George, R., Gonzalez, M., Guartin, H., Harris, N., Li, P.,									
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuno, J., Paclob, J.,		Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuno, J., Paclob, J.,									
Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,		Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,									
Celniker, S. and Rubin, G. M.		Celniker, S. and Rubin, G. M.									
BDCP/HMT AT Drosophila EST Project		BDCP/HMT AT Drosophila EST Project									
Unpublished (2000)		Unpublished (2000)									
On Dec 6, 2000 this sequence version replaced gi:11574240.		On Dec 6, 2000 this sequence version replaced gi:11574240.									
Contact: Stapleton, M.		Contact: Stapleton, M.									
BDCP		BDCP									
Lawrence Berkeley National Lab		Lawrence Berkeley National Lab									
One Cyclotron Rd Berkeley, CA 94720, USA		One Cyclotron Rd Berkeley, CA 94720, USA									
Fax: 510 486 6796		Fax: 510 486 6796									

## ALIGNMENTS

AT1884725	w183h06.x	AT1884725	w183h06.x
AT667593	AL667593	AT667593	AL667593
AU002272	AU002272	AU002272	AU002272
AL611639	Anopheles	AL611639	Anopheles
AT293077	CH16161.5	AT293077	CH16161.5
BM653918	17006873	BM653918	17006873
BM581199	17006872	BM581199	17006872
BM693831	UI-M-EGO-	BM693831	UI-M-EGO-
AL157191	Anopheles	AL157191	Anopheles
BC033610	Mus muscu	BC033610	Mus muscu
AL064294	Drosophila	AL064294	Drosophila
BM639251	17006875	BM639251	17006875
-AJ282560	4A3A-AAZ-	-AJ282560	4A3A-AAZ-
BM603942	17006590	BM603942	17006590
BO885751	AGENCOURT	BO885751	AGENCOURT
BP013653	BP013653	BP013653	BP013653
AK014572	Mus muscu	AK014572	Mus muscu
BG782849	SEAMC002	BG782849	SEAMC002
AV884082	AV884082	AV884082	AV884082
AV904264	AV904264	AV904264	AV904264
BF491209	AT27901.5	BF491209	AT27901.5
BT155965	602904274	BT155965	602904274
AV904484	AV904484	AV904484	AV904484
BQ425264	AGENCOURT	BQ425264	AGENCOURT
AK018553	Mus muscu	AK018553	Mus muscu
AK014544	Mus muscu	AK014544	Mus muscu
AT064231	GH04562.5	AT064231	GH04562.5
BM588494	170006873	BM588494	170006873
BT729843	603349502	BT729843	603349502
BM951058	UI-M-EGO-	BM951058	UI-M-EGO-
BE753691	206827 MA	BE753691	206827 MA
BM691900	CH1A4813T	BM691900	CH1A4813T
AL100001	Drosophila	AL100001	Drosophila
BO941298	AGENCOURT	BO941298	AGENCOURT
BT240479	RE37330.5	BT240479	RE37330.5
AU131521	AU131521	AU131521	AU131521
BM644904	17006873	BM644904	17006873
N34242	y*79c08.r1	N34242	y*79c08.r1
BM595852	17006874	BM595852	17006874

Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic AE003832: arm:2R [4615216..4876965]  
 estimated-cyto:45F4-46B13: 04/09/2001  
 Plate: AT.275 row: G column: 3  
 High quality sequence stop: 637.

## FEATURES

source

1. 682  
 Location/Qualifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_1lb="AT27575"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates AT.10-AT.120: DH5-alpha. Plates

AT.121-AT.319: DH5-alpha Tona"

/note="Organ: ADULT testes. Vector: pOTB7; Site\_1: EcorI;

Site\_2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."

BASE COUNT 120 a 214 c 196 g 152 t  
 ORIGIN

## Query Match

Best Local Similarity 7.0%; Score 200.8; DB 12; Length 682;

Matches 310; Conservativeness 0; Mismatches 182; Indels 0; Gaps 0;

684 GTACGACGAGTCTCACTACTTCTTTCAGAACAGTTCTCCAAACAGGAGGAAATG 743  
 191 GAACCAACCTCGGGGAGTTCTATTCAGAACATCTTCGCGAGAAAGTACGATTCG 250  
 744 AAGAGAGTCTCGGGGAGTTCTATTCAGAACATCTTCGCGAGAAAGTACGATTCG 803  
 251 ATGATGACATCGGATATCCAGCTGAGCTGGCGCTGGAGCTGGCTGGATTCG 310  
 804 TGGTGTTCGAGTCTCGCCGCGAGAGTCAAGAGTTCCGCAAGCGGCTACTTCCTCG 863  
 311 TCATCCGCGGGGATATGTTCAAGGTTGTAAGAGTTGCGGCAAGGCTCTACTTCCTCG 370  
 864 CGCTCTCCCATACGTTGTATGATGATCACTTATTCATCAACATCATCTCCCGGTG 923  
 371 CCTCTCTCCCATACGTTGTATGATGATCACTTATTCATCAACATCATCTCCCGGTG 983  
 924 CTATGACGAGCATCTGTTCTGTCACGCTCAATGAGGAGCAATCTGAGCTGGTG 983  
 431 CCTTCAGCGGCTTCTGTTCTGTCACGCTCAATGAGGAGCAATCTGAGCTGGTG 983  
 984 TATGATCTCAGACGATCGCAAGAGTCTTCTCTGACAGTGTGACCGGACCGATCA 1043  
 491 TCTGTACGCGCGCGGCGGCGGAGTCTTCTCTGACAGTGTGACCGGACCGATCA 550  
 1044 TCATGTTCTCTCTCAAGGTTTCAGACATATATCTACAGGAGTGTGATGTGTA 1103  
 551 TCATGTACGCTCTGTACACCGCTTCGCGGCAACATTTACAGGAGTGTGATGTGTA 610  
 1104 CGACTTTGACACCTTTACAAAGTTTCTGCTGAGTGTGAGTGTGATGTGTA 1163  
 611 CCAGCGTGTGACACCTTCACTCTCTGAGTGTGAGTGTGATGTGTA 670  
 1164 ACCTCGCGTACG 1175  
 671 ACCTGCGGTACG 682

## RESULT 2

BM589455

LOCUS

DEFINITION 704 bp mRNA linear EST 25-FEB-2002

19600449705511 5', mRNA sequence.

ACCESSION

BM589455

VERSION

BM589455.1

KEYWORDS

EST.

SOURCE African malaria mosquito.

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

## REFERENCE

1 (bases 1 to 704)

Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlalab

R., Collins, F.H., Venter, J.C. and Hoffman, S.L.

Unpublished (2002)

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404531351

Fax: 2404534580

Email: [holtra@celera.com](mailto:holtra@celera.com)

Plate: N001004A80 row: 0 column: 01

Seq primer: M13 Reverse.

## FEATURES

source

1. 704  
 Location/Qualifiers

/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std.

chromosome)"

/db\_xref="taxon:7165"

/clone\_1lb="19600449705511"

/dev\_stage="Adult"

/lab\_host="DHI10b"

/note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'.

Clones available through the Malaria Research and Reference Reagent Resource Center ([www.malaria.mrl.org](http://www.malaria.mrl.org))"

BASE COUNT 131 a 220 c 185 g 168 t  
 ORIGIN

## Query Match

Best Local Similarity 5.7%; Score 163.6; DB 13; Length 704;

Matches 359; Conservativeness 0; Mismatches 304; Indels 3; Gaps 1;

223 CACGAGCTTAGAGGCTGACCGGCGAGAGTATGATGTGTCACAAACATGTAATCTT 282  
 29 CACCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 88  
 283 GATGCTCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 342  
 89 ACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 148  
 343 GTACCAAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 402  
 149 GTTCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 208  
 403 CAAGCCTGTGTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 462  
 209 ACGCCGATCTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 268  
 463 AGTTGGTCAATTTACCGGCGCATGAAAGTATGATAGTATGATAGTATGATAGTATG 522  
 269 GGTGTAGACACTTACCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 328  
 523 TTACATCTGCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 582  
 329 CACCGTATCTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 388  
 583 CTTCAGAGGCACTCTTCATGAGGCTATTGTTCAGCCTGAGTGGAGAACTGCTACCTTC 642  
 389 GTTCAATCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448  
 643 AGATTCACACTT---GCTGCTACAGTCAACATCAACATCAACATCAACATCAACATCA 659  
 449 TACCGGCTCTGAGTATGCTGCTGAGTATGCTGAGTATGCTGAGTATGCTGAGTATG 508

QY 700 ACTGACTTTTGAGAACAGTTCCTCCACAAAGGATGGAATGAGAGGCTCGGTGC 759  
 Db 509 TCTGTACTCCGCGAAAGATGATGCGCGTACACTGACATGAGATGGTTGGGATA 568  
 QY 760 CCCATCTGTGACTGCTGTGTCTATATCATGTCATGCTCATGGTTCGAGATCGT 819  
 Db 569 TCCGAGTGGGCGCTGACATGCTGATGCTGCTGCTGATGCTGATGCTGATGCTG 628  
 QY 820 CGCCGAGAGTCAAGAGTCCGCGAAGGCGCTACTCTCCGCGCTTCGCATACGT 879  
 Db 629 GATCAAGGATTAAGAGCTCCGCGAAGGCTTCTACTCTTCCATCTTCCGATAC 688  
 QY 880 TGTGAT 885  
 Db 689 CATCAT 694

RESULT 3  
 CNS0110L/777 bp DNA linear GSS 14-JUN-2001  
 LOCUS  
 DEFINITION Anopheles gambiae GSS SP6 end of clone 19M22 of Notre-Dame library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.

ACCESSION AL149894.1 GI:7010373  
 VERSION  
 KEYWORDS African malaria mosquito.  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

REFERENCE 1 (bases 1 to 777)  
 AUTHORS Direct Submission  
 JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage; BP 191 91006 Evry cedex - FRANCE (E-mail: sequef@genoscope.cns.fr)  
 Web: www.genoscope.cns.fr

REFERENCE 2 (bases 1 to 777)  
 AUTHORS Roth,C.W., Brey,P.F., Ke,Z., Collins,F.H. and Weissenbach,J.  
 JOURNAL Direct Submission  
 Submitted (16-FEB-2000) BIMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France

COMMENT This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.

FEATURES  
 source location/Qualifiers  
 1..777

/organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="19M22"  
 /clone\_1lb="Notredame1"  
 /note="end : SP6"  
 BASE COUNT 194 a 209 c 217 g 145 t 12 others  
 ORIGIN

Query Match 5.3%; Score 151.6; DB 17; Length 777;  
 Best Local Similarity 54.8%; Pred. No. 8.9e-17;

Matches 311; Conservative 5; Mismatches 249; Indels 3; Gaps 1;

QY 740 ATTGAGAGCTTCGGTCCCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 799  
 Db 743 ACTGCGGCTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 684  
 QY 800 CTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 859  
 Db 683 GTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 624  
 QY 860 CTCGCGCTTCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 919  
 Db 623 CTGGCCCTCTTCCCGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564

QY 920 GGTCTACTGACGCGCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 979  
 Db 563 GCGCCCGCTGACGGCTGTAAGTACTTATCATACACCCGATCGGATCACTTAATCG 504  
 QY 980 GGTGTATGCTACGACAGTACGACGACGACGACGACGACGACGACGACGACGAC 1039  
 Db 503 GACCTGAGAGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444  
 QY 1040 ATCATCATGCTCTCTCTCTTAAACGCTTACGACATATATATCTACAGGATGCT 1099  
 Db 443 GTCAATGCTCTCTCTCTTAAACGCTTACGACATATATATCTACAGGATGCT 384  
 QY 1100 GTTACGACTTGTGACACCTTATACAGTCTTCTGCTGCTGCTGCTGCTGCTGCT 1159  
 Db 383 ATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 324  
 QY 1160 GGTACCTGCGCTACGACCTACCTACGACGACGACGACGACGACGACGACGAC 1219  
 Db 323 GCGCATCTGCGACATCTGACAAACGACCTACACCAAGGTGTAAGCCGCGCAGT 264  
 QY 1220 AGCTTGTGCTTATTTATACCTGACGACGACGACGACGACGACGACGACGAC 1276  
 Db 263 GCGCTAACGTTATACCTACCTACCGACGACGACGACGACGACGACGACGACGAC 204  
 QY 1277 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1304  
 Db 203 TTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 176

RESULT 4  
 BC026339 3118 bp mRNA linear HTC 08-APR-2002  
 LOCUS  
 DEFINITION Homo sapiens, soluble carrier family 6 (neurotransmitter transporter, betaeline/GABA), member 12, clone IMAGE:479451, mRNA.  
 ACCESSION BC026339 GI:20072875  
 VERSION  
 KEYWORDS HTC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 3118)  
 AUTHORS Strausberg,R.  
 JOURNAL Direct Submission  
 Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)

Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdickpa@stanford.edu](mailto:mdickpa@stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
 Series: IRAK plate: 32 Row: a Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507036  
 this clone has the following problem: frame shifted.

FEATURES  
 source location/Qualifiers  
 1..3118

/organism="Homo sapiens"  
 /db\_xref="LocustID:5539"  
 /db\_xref="taxon:9606"

```

/clone="IMAGE:4794351"
/issue-type="Brain, hypothalamus"
/ld_host="NIH_MGC_96"
/ld_host="DH10B"
/ld_host="Vector: Bluescript"
BASE COUNT      636 a      885 c      832 g      765 t
ORIGIN

```

Query Match 5.28; Score 149.6; DB 11; Length 3118;

Best Local Similarity 46.38; Pred. No. 1.1e-16; Mismatches 814; Indels 37; Gaps 6;

```

Matches 735; Conservative 0; Mismatches 814; Indels 37; Gaps 6;

221 GACAGGACTTAGAGGCTGACCGCCAGAACGATAGGTATGTCACAAACATGTAATTC 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GACAGGAAACGAGGACGAGGATGAGGATCGGGGCCAATGAGCAACAAAGATGATTT 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 281 TTGATGCTTCATCGCTACATCCGTCGTTGGGTAGACGTGGGGGTTCCCTTTCATC 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 GTCTGTCACTGGCCGGGAGATCATTTGGGCTGGCAATCTGAGAGTTTCCCTATCTC 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 341 GCGACAGATAGGAGGAGTGTCTTCTGTCAGTACAGTCAATCGTTCTTACTGTC 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 TGTACAAACGAGGATGAGGCTTCTTATCCCTACTTCACTTCTTCTTGTCTGC 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 401 GGCAGGCTGTACTACTAGAGTGTCTCGACAAATTCAGTTCAAGAACTCTGT 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 GGCATCCCGGTCTTCTGAGAGGTGGCGTGGGCCAATACACAGCCAAAGGAGTCTC 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 461 AAGATT---GTCATTTACCGGGCCATAAAGTACTGATACGCTCAAGCTGCCGC 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 ACAGCCGAGAGAGATCCGCCCTTCCAGGGCAATGCTGCGCATCTGTGCTATC 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 518 TGGGCTTACATCTGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 GAGTCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 578 ATGAGCTTCAGGCGCTTCATGAGGATTTGTCAGGCTGAGTGGAGAACTGGGTA 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 AGTCCTCACTTCTGAGCTGCCCTGGACCCGCAACAACTTTGGAACACAGACAT 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 638 CCTCAGAT-----CCAACACTGTGATCACTCAACCAACATCACCATTGATACC 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 TGCAGGACTTTTGAACCACTCAGAGCCGCGACATGACCCCATTTGAAATTTAAC 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 689 AGCAGGCTCACTCTACTTTTGAAGAACATTTCTCCAAACAAAGCGATGAAATGA 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TCACCTGTATGAAATCTGAGAGACGAGTTCTGGCATACCTGGGGATCA---T 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 749 GGTCTGGTCCCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 GACCTGGGCTCCCTGCGCTGGAGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTG 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 809 TTGCGAGTCTGCGCCGAGAGTCAAGATTTCCGCGCAAGCGGCTTCTCTCGGCTC 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 TATTTCTGATCTGGAAGGGGCTCAATCCACAGGCAAGGTGTTTATTTCAACGCC 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 869 TTCCCATAGCTGTATGATCACTTATTCATCAACCAATCACTCTGCCGCTGCTACT 928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1009 TTTCCTGATGATGCTTGTATTTGCTGATGAGAGTGTCAACCCCTCCCGAGCTAC 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 929 GAGGCTCTCTGCTTCTGCTGAGCCCTCAATGAGCCCAACTCTTGAAGCTGCTATG 988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1069 CAGGGCACTCTACTACTTGAAGCCAGATTTTCCGCTCAAGAGCCCTCAGAGTCTG 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 989 TACTCAGCATCTGAGCAAGTGTCTTCTGCTGACAGTGTGACCCGAGCATCATG 1048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1129 ATGAGATGCGGACCCGATCTTCTTCTTGTCCATCTGCGAGGGGCTCCCTGACAGCC 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1049 TTCTCTCTTCAACAGGTTTCAAGCATATATCTACAGAGATGCTTGATTTGATGACT 1108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1189 CTGGGAGCTACAAACAGATCAACAACTGCTACAAAGAGCATGCGCCCTGTGCTC 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1109 TTGGACACCTTTCACAGTTCTGTCTGTGGTGCACGATCTTGGGATCTTGTACCTC 1168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1249 CTGACAGTGCACACACACTTTGAGCGTGGGCTTTGTGTCTC-----TCCATCTGGGGCTTC 1304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1169 GCGTAGCACTCAACTCAGAGGTGGAGATGTGTGCTGCTGCGGTACACAGTCTGCT 1228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1305 ATGTCCCAAGAGAGAGGGGTGCCATTTCTGAAGTGTGCCAGTACGCTCTGGGCTGGCC 1364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1229 TTCAATTCATACCTGATGCTGATTCGCAAAACATTCACACCTGATCTCTGGGCTG 1288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1365 TTCTATGCTTCCCAAGGCTGTGACTATGATGCTCTTATCCAGCTGTGTGCTGCTG 1424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1289 TTCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1425 TTCTTTATGATGCTATATTCAGAGGCTGACAGCAGATTTGCTGTGAGTGTGCTG 1484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1349 AACACATGGCGATGACGCGTTC-----AGTGTACCCACCGTCTAC 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1485 GTGACACCTTCACATAGATGTTCCCGAGCAGCTCCGGAAGACGGGCGCGAGCTC 1544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1394 ATGTACAGCATGACCTGCTTCTGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1545 CTATCTCACCATCGCCGCTGATGCTACTGATGAGGCTTTTCTGTGACCGAGGCG 1604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1454 GACAAATATATCTTGAAGCTGTAGATCACTACGCTGGAACATTCCT---TGTCTTTC 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1605 GGAATGATATCTTCCAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1511 TGGGCTTCTGATACCTGACGCGTGTCTGATTTATGATTTGAGAAATCTGTGCTA 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1665 CTGATGATTTGTAAGGAGTGTGATTAAGTGTGATGATGATGATGATGATGATGAT 1724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1571 GACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1725 AACATTTAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1631 ATCAGCTCTCTATATGACAGCTGTGTCTTCTGCTCTTCTGCTTCAACACCTG 1690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1785 CTGACCTCTGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1691 GTGTGGAGACAACTGATTAATCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1845 AAGTACAAACAGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1751 TTTTGGGCACTGACGTTTGTGCCAAT 1776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1905 CTGTCTCCATGATGCTGTGTGCCACT 1930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5  
Bg780471 798 bp mRNA linear EST 20-MAY-2001

LOCUS  
SEAM0000428 Sea urchin primary mesenchyme cell cDNA library  
DEFINITION  
Strongylocentrotus purpuratus cDNA clone  
91222851.N20\_072\_PC\_0020\_B2\_MR\_610 5', mRNA sequence.

ACCESSION  
Bg780471  
Bg780471.1 GI:14151484  
VERSION  
EST.  
SOURCE  
Strongylocentrotus purpuratus.  
ORGANISM  
Strongylocentrotus purpuratus

REFERENCE  
Zhu,X., Mahlras,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and  
Ettensohn,C.A.  
A large scale analysis of mRNAs expressed by primary mesenchyme  
cells of the sea urchin embryo  
Development 128 (13), 2615-2627 (2001)

JOURNAL  
MEDLINE  
CONTACT: Ettensohn CA  
Dept. Biol. Sci.  
Carnegie Mellon University  
4400 Fifth Avenue, Pittsburgh, PA 15213, USA



[illegible]

Db		602	TGCGTCGGTGTCTCNTGTTGGTGTGTGTCTCCTCCAGCCATCCGAGGGGTCAAGTCTTCA	543
OY		842	GCGAAGAAGGCCTTAATTCTCTCGCGCTCTTCCCATACGTTGTGATGATCACATTATCATC	901
Db		542	GGGAAGATGTGTACTTTCACGCCACAGCTTCCCACAGTGAGTGACCATTTCTTTGTTC	483
OY		902	ACCACAAATCATCCTCGCCGGTGTACTAGACGGCAATCCGTTCTTGTCACGCCCTCAATGG	961
Db		482	CGCCGAGTGAACCTCTGGAGGGAGCCTTTGATGACATCATGTACTACCTAACCCCGCAGTGG	423
OY		962	GCGAAGTCTCTTGAAGCTGGTGTATGTGTACTACAGCATGCAGCAAGTGTCTTCTCTGTG	1021
Db		422	GACAAGATCTCGTAGAGGCCAAGGTGGGGTGTATCTGCTGCCAGATCTTCTACTCACTG	363
OY		1022	ACAATGTGACACCGACCGATCATATGTCTTCCTCTTCAACAGGGTTTCAGACATAATATC	1081
Db		362	GCGTCGCGGTGGGAGGCTTCATACATGAGCTTCTCTCAACAAATTCACAAATTAATCTGT	303
OY		1082	TACAGGAGATGCTTGATGTGTATACGACTTTTGGACACCTTTACAAAGTTCTTGTCTGGGTGC	1141
Db		302	TACCGGGACAAATGTATCATATCATAGATATACCAAATCTGTGGCACAGCGCTATAGCTGGCTTC	243
OY		1142	ACGATCTTCGATTCCTTTGGTATACCTCGCGTACGAACTCAAGTCAAGGTGGGAGATGTG	1201
Db		242	GTCATCTTCTTCACATCCCGCTGCTTATGCGCAATACCTCGGCGCTGAGTGTCTCCGCTGTG	183
OY		1202	GTCGCTGCTGGCGGTACACAGTCTTGCTTTCATTCATACCTGTAGTCCATGTCCAAAACA	1261
Db		182	---GCAGACACAGGCCCTCGGCTGGCTTCTGTGGCTTACCCCCGAGGCCCTCACACTACTT	126
OY		1262	TTCCAACTCATGCTATTTCTCGGTCTGTCTTCTCTCATGATGTGCGTGTGGATGTGGCG	1321
Db		125	CCCATCTTCCCGCTGTGTCT	66
OY		1322	TCATCCGCGGCTGTGCTATGCACTTTCAACCAATTTGGAGATGGAGCGGTT	1371
Db		65	ACTAGTTTCTGCTCTCTCTGGAGACCTGTGTACACACCATTTGTGATGAGT	16
RESULT 8				
LOCUS		AL667593	846 bp	mRNA linear EST 11-JAN-2002
DEFINITION		AL667593 directional larval cDNA library Clona intestinalis cDNA		
ACCESSION		AL667593		
VERSION		AL667593.1	GI:18134500	
KEYWORDS		EST.		
SOURCE				
ORGANISM				
				Clona intestinalis.
				Clona intestinalis.
				Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
				Phlebobranchia; Clonidae; Clona.
REFERENCE				1 (bases 1 to 846)
AUTHORS				Genoscope.
TITLE				Clona intestinalis directional larval cDNA library
JOURNAL				Unpublished (2002)
COMMENT				Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr IMPORTANT: this sequence may contain errors. The Clona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya. Directional larval cDNA library originate from Dr.M.Brammo, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescriptSK+.
FEATURES				
source				location/Qualifiers 1..846 /organism="Clona intestinalis" /db_xref="taxon:7719" /clone="021ZF11" /clone_lib="directional larval cDNA library" /note="Vector: pBluescriptSK+" BASE COUNT 199 a 171 c 197 g 277 t 2 others





## COMMENT

This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.

## FEATURES

Location/Qualifiers  
1. 527

/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="31F21"  
/note="Notredame1"  
/note="end : SP6"

BASE COUNT 90 a 157 c 138 g 142 t  
ORIGIN

Query Match 4.6%; Score 132.4; DB 17; Length 527;  
Best Local Similarity 56.7%; Pred. No. 2.3e-13;

Matches 244; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 740 ATTGAAGAGGTCTCGGTCGCCCATCTGCTACTGTGTGTCTATTATCATGCATGG 799

Db 84 ACTGCGGGTGTGCTGGTGTGGGACTGGAGCTGACGCTGCTGCTCTTCATCTGG 143

QY 800 CTATAGGTGTTCGAGTCTGCGCCGAGAGTCAAGATTCCGGCAAGCGGCTACTTC 859

Db 144 GTGGCGATGACATCATGATGTGAAGGATTCGCGGATCCGCAAGTCTGATCTTT 203

QY 860 CTCGCGCTTCCCATAGTGTTCATGATCATCTTATTTCACCAATCATCTGCCCC 919

Db 204 CTGGCCCTCTCCGATGCTGTGCTGATCTCTGTCGATCTTCAGCTTACAGCTGAG 263

QY 920 GGTGCTACTGACGACATCTGCTTCTGTCACGCTCAATGGCGAACTCTTGAGCTC 979

Db 264 GCGCGCGGTGACGAGCTTGAAGTACTTCATCATCCCCCAGCTGATCGTGAATGCG 323

QY 980 GGTATATGATCAGACGATGATCTCTCTCTGACAGTGTGACCGGACCG 1039

Db 324 GACGCTGGAAGAGGCGCTGCTGCAATGTTCTCTGCTGCTGATCTTTGGGCGGT 383

QY 1040 ATCATCATGTTCTCTCTTCAACAGGTTTCAGACATTAATCTACAGGATGCTTGAT 1099

Db 384 GTCATTCCTCTCTCTCTTCAACAACTTCAGACAAATTTACCGGATGAGATGATC 443

QY 1100 GTTACGACTTGGACACTTACAGTTTCTTGTCTGGTGACAGATCTTCGATCTT 1159

Db 444 ATCTCGGCTGATACCTTCACTCTCTGCTGCGAGCGCTGCTCTCTCATCAT 503

QY 1160 GTTACCTCG 1169

Db 504 GCGCATCTGG 513

RESULT 11 811 bp mRNA linear EST 23-Apr-2001  
AI293077

LOCUS GH16161.5prime GH Drosophila melanogaster head port2 Drosophila

DEFINITION melanogaster cDNA clone GH16161 5 similar to BG:DS03431.1:

FBan0015279 located on: 2L 35B7-35B7; 04/10/2001, mRNA sequence.

ACCESSION AI293077 GI:13759911

VERSION EST.

KEYWORDS fruit fly.

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 811)  
Harvey, D., Broksstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S. and Rubin, G. M.  
BDGP/HMI Drosophila EST Project  
Unpublished (2001)  
On Dec 1, 1998 this sequence version replaced gi:3942484.  
Other ESTs: GH16161.3prime

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA.94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

hit genomic AE003645: arm:2L [14622887,14883853]

estimated-cyto:35B5-35B9: 04/10/2001

plate: GH.161 row: F column: 1

High quality sequence stop: 805

POLY-A-NO.

## FEATURES

Location/Qualifiers  
1. 811

/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="GH16161"  
/clone\_lib="GH Drosophila melanogaster head port2"  
/sex="male and female"  
/dev\_stage="adult"  
/lab\_host="DHS - alpha"  
/note="Organ: head; Vector: port2; Site: 1; EcorI: Site: 2;  
XhoI: sized fractionated cDNAs were directly ligated into  
port2. Plasmid cDNA library."

BASE COUNT 196 a 199 c 214 g 202 t  
ORIGIN

Query Match 4.6%; Score 130.8; DB 9; Length 811;  
Best Local Similarity 53.9%; Pred. No. 3.6e-13;

Matches 293; Conservative 0; Mismatches 247; Indels 4; Gaps 1;

QY 88 TATTGTGACAAATGATGATGACGCGCAAGTAAACGGGGTTTCAGTGTGAGCCAA 147

Db 27 TACACTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 86

QY 148 GATGGAACCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 207

Db 87 GAGGAGCGAGTACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 142

QY 208 GATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 267

Db 143 AGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 202

QY 268 CAACATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 327

Db 203 GGGGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 262

QY 328 GTTCCCTTTCATGCGGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 387

Db 263 ATTCCCTTTCATGCGGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 322

QY 388 TCTTTTACTTGTGCGGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 447

Db 323 CTGTTCCTGATGCGGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 382

QY 448 AAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 507

Db 383 TCGGGGCTCGTTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 442

QY 508 AGCTGCGGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 567

Db 443 GCGCTTTCAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 502

QY 568 TTACTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627

Db 503 ATATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 562

QY 628 GAAC 631

Db 563 TATC 566

RESULT 12  
BM653918



LOCUS BM653918 707 bp mRNA linear EST 26-FEB-2002  
DEFINITION 17000687379138 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone  
ACCESSION 19600449619090 5', mRNA sequence.  
VERSION BM653918  
KEYWORDS BM653918.1 GI:18953429  
SOURCE EST.  
ORGANISM African malaria mosquito.  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
REFERENCE 1 (bases 1 to 707)  
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltR@celera.com  
Plate: N001004NBT row: N column: 16  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..707  
/organism="Anopheles gambiae"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449619090"  
/clone\_lib="A.Gam.ad.cDNA1"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."  
BASE COUNT 141 a 187 c 189 g 190 t  
ORIGIN  
Query Match 4.4%; Score 126.4; DB 13; Length 707;  
Best Local Similarity 59.4%; Pred. No. 2.2e-12;  
Matches 214; Conservative 0; Mismatches 146; Indels 0; Gaps 0;  
QY 739 AATTGAAGAGGCTTCGCTGCCCTCCCATCTGTTGTTGTTCTATTTCATCGCATG 798  
DB 335 ATTGCGAGAGGAGCTTGATGATGCGAGATGCGACAGCTGCTCTGCTACTGCTCTCG 394  
QY 799 GCTCATGGTGTGGGAGCTGCTGCCCGGAGTCAAGGTTCCGGCAAGGGGCTACTT 858  
DB 395 GTCCATCTGCTTGTGATCTCATCAAGAGATGCGACCGGGAAAGGGGCTACTT 454  
QY 859 CCTGCGGCTCTTCCCATACGTTTCATGATGATCTTATTCACCAACATCATCTGCC 918  
DB 455 TCTTGCCATCTTCCCGATGATATTCATCTCTGCTTCCACACAGTCCGACGTCG 514  
QY 919 CGGTGCTACTGAGGAGCTCTGTTCTTCTGTCAGGCTTCATGCGGAAATCTCTGAGCT 978  
DB 515 AGGTTCCTCGAGGATATCAAGTTCTTCTTAACCCCAATGGGAAAGCCATATCTCAGC 574  
QY 979 CGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038  
DB 575 CAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 634  
QY 1039 GATCATCATGTTCTCTTCAACAGGTTTCAGATATATCTACAGGATGCTGATGAT 1098  
DB 635 AATCATGCTAGTCTCTGTTCAATATTTCTCAATATATGATGATGATGATGATGATGAT 694

RESULT 13

LOCUS BM581199 589 bp mRNA linear EST 22-FEB-2002  
DEFINITION 17000687244048 A.Gam.ad.cDNA blood1 Anopheles gambiae cDNA clone  
ACCESSION 19600449683580 5', mRNA sequence.  
VERSION BM581199  
KEYWORDS BM581199.1 GI:18896666  
SOURCE EST.  
ORGANISM African malaria mosquito.  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
REFERENCE 1 (bases 1 to 589)  
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltR@celera.com  
Plate: N001004B2V row: M column: 06  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..589  
/organism="Anopheles gambiae"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449683580"  
/clone\_lib="A.Gam.ad.cDNA.blood1"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not 1 site is 3'.  
Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)."  
BASE COUNT 106 a 173 c 150 g 160 t  
ORIGIN  
Query Match 4.4%; Score 124.8; DB 13; Length 589;  
Best Local Similarity 52.3%; Pred. No. 4.5e-12;  
Matches 301; Conservative 0; Mismatches 272; Indels 3; Gaps 1;  
QY 1030 CACCGAGCCGATCATCATGTTCTCTCTTACACAGGTTTCAGACATATATCTACAGGGA 1089  
DB 3 CACGGCTCGGCTGATATGATATGCTATGCTCTTACAGACTTTCACAAACATCTACCGGCA 62  
QY 1090 TGCCTTGATTTTACAGACTTTTGACACCTTTACAGAGTTTCTGTTCTGGGTGACAGATCTT 1149  
DB 63 TGCCTGATATCATCACTCATCTACGACATGCTACTTGGTGGCGGGTCTGTTGCTT 122  
QY 1150 CGGTATCTCTGTTACCTCGGCTAGCACTCACTACAGAGGAGGAGATGTTGTCGGTGC 1209  
DB 123 CGGTGATTTGATCTACTAGCTACGCTACCGGTCAGTCCGACATCTCCAAAGTGTCCA 182  
QY 1210 TGGCGGTACCACTCTGCTTCAATTCATACCTGATGCCAATGCGAA--ACATTCGA 1266  
DB 183 AAGTGGCCACACCTTACTTCACTACCTACCGGACAGATGCGAAGTTGACTTCT 242  
QY 1267 ACCTGAGTATTTCTGCTGCTGCTTCTCTTATGATGTCGCTGGGTATCGCTATC 1326  
DB 243 GCCGCAATTTCTTCGCGGATTCCTTCTTATGCTGTTCTCGGCTGCGCATGTTGACCT 302  
QY 1327 CGTGGCTCTGCTATGCACTTTCACACATGTCGATGAGGAGGATGCTCCAGCTGATACCCAC 1386  
DB 303 GATGGGATGCTATGATGCTGATATGATGATGATGATGATGATGATGATGATGATGAT 362  
QY 1387 GGTCTACATGTCAGGATGACCTGTTCTTGGCGTTCTGCTGTTGATGTTGTTACTGAC 1446

Db	363	ATGGAAGATCGTATCGTCGGCGGCTAGCAGGGTTCGCAFTTGGGTTGGCTTACATAC	422		
Qy	1447	ACCGGCTGCACATATATCTTGGACCTTGTAGATCATTACGAGTGGACATTCCTTGTCT	1506		
Db	423	ACCGGCTGTTTGAATCATCTAGCTAGCACTCTGATATTACAGCGCACGCTGGTAAACAT	482		
Qy	1507	TTTTCGCCCAATTTCTGAATCGCAGGCGGTGTCTGGATTATGATTTAGTGGAAATCTGT	1566		
Db	483	AACGTGGCCGCTCTTCGAGCTGCTCACCCTTTGCTGGATTCATATGCGCTGAATCGGCTGT	542		
Qy	1567	CCTAGACATTGAGTTCATGTTGGTAAAGACTGG	1602		
Db	543	CAGGACATGTGATTCATGCTCGGCATCAAGACAG	578		
RESULT 14					
LOCUS	BM963831	777 bp	Linear EST 18-MAR-2002		
DEFINITION	UI-M-E00-bwn-j-09-0-UI.r1 NIH_BMAP_E00 Mus musculus cDNA clone				
ACCESSION	BM963831	IMAGE:5698664 5', mRNA sequence.			
VERSION	BM963831.1	GI:19547251			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 777)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgsab@remail.nih.gov				
	Tissue Procurement: Dr. James Lin, University of Iowa				
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa				
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa				
	Clone Distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/ILMIL at:				
	http://image.llnl.gov				
	This clone was contributed by the Brain Molecular Anatomy Project				
	(BMAP)				
	Seq primer: pyx-5.				
FEATURES	Location/Qualifiers				
Source	1..777				
	/organism="Mus musculus"				
	/strain="C57BL/6"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:5698664"				
	/clone_1lb="NIH_BMAP_E00"				
	/issue_type="Whole Brain"				
	/dev_stage="embryo 15.5 dpc"				
	/lab_host="DHL0B (T1 phage resistant)"				
	/note="Organ: brain; Vector: pyx-Asc; Site: 1: Ecor I; Site-2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I, and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System", supported by National Institutes of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."				
BASE COUNT	134 a	233 c	193 g 215 t 2 others		
ORIGIN	Query Match				
	4.2% Score 121.4; DB 14; Length 777;				

	Best Local Similarity	50.8%	Pred. No. 1.6e-11	
	Matches 370;	Conservative 0;	Mismatches 346;	Indels 11; Gaps 3;
QY	766	CTGTAATCTGTGTGTGTGTATTCATTCATGCAATGGCTCAATGGTGTGGAGTCTGCGCCG	825	
Db	24	CTGGCCTCTGGGCATCACACTGGCATTCGCTGGGTGCTGTGATATTTCTGCATCTGGAA	83	
QY	826	AGGAGTCAAGAGTTCCGGCAAAAGGGGCTACTTCTCGCGCTCTTCCCATACGTTTCAT	885	
Db	84	GGGTGTGGTGGAGTGGAAAGGTGGTCTACTTCTTCACCCAGCAATGATCATCATGCT	143	
QY	886	GATCACTTATTCATCATCACCAATCATCTCTCGCGTCTACTGACGAGCATCTGTCTT	945	
Db	144	TATCATCTGTCTTCTCGTGGAGTACGACCTTCCGGGGCCAAAGAGGGAGATCTCTTCTA	203	
QY	946	CGTCAAGCCTCAATGGCGGAAATCTTGTAGCTCGGTATATGTACTCAGACGTACGCA	1005	
Db	204	CATCAACACCCAAATCTCCGAAAGCTGTCTGATTCATGAGGTGTGGCTTGACGGCCACCCA	263	
QY	1006	AGTGTCTTCTCTCTGACAGTGTGACGGGACGGGACCATCATATGTTCTCTTTCAACCG	1065	
Db	264	GATCTTCTTCTCTCTCTGACGGGCTGTGGCTGTGGCTTCTGTCTGTGGAGTCTCAATCT	323	
QY	1066	TTTCAGACATATATCTACAGAGATGCTTGGATTTTACGACTTTGACACCTTTTACAG	1125	
Db	324	TTTCCACAAACATGTGTACAGGACCTCATCATGTGTGTGATTCATCACTCTCCACAG	383	
QY	1126	TTTCTGTGTGGGTGCACGATCTTCGATCTGTGTACTGTGTAACCTCGCATGCAATCTCA	1185	
Db	384	CATGTTTCCGGATTCGATCT	443	
QY	1186	AGAGTGGAGATGT	1245	
Db	444	GTCCATACATGATGT	500	
QY	1246	TGCCATTTGCCAAAACATTTCAACTCACTCACTATTTCTCGTGTCTGTCTTCTGTATGTC	1305	
Db	501	GGCTGTGCACAGGTACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	560	
QY	1306	GGTGTGTGGATTCGCTGT	1365	
Db	561	GATCTGTGGGATTTGACACCCAGTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	620	
QY	1366	CGCGTCTCCACGTGTACCCACCGTCTACATG-----TCAGCGATGACCTGTCTTGGCG	1419	
Db	621	CGAGTACCCCAAGCTTCTCCGCAATGCGCGTGAATCTTCAATGCTTTCAGTGTGATGCTG	680	
QY	1420	T--TTCTGT	1477	
Db	681	TCCTACATGATGT	740	
QY	1478	GATCACTAC 1486		
Db	741	GATATTTAC 749		
RESULT 15				
CNS01RBB		584 bp	DNA	linear
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				

## JOURNAL

Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail) : [segrefigenoscope.cns.fr](mailto:segrefigenoscope.cns.fr)  
- web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
3 (pages 1 to 504)

## REFERENCE

AUTHORS  
TITLE  
JOURNAL

Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.  
Direct Submission  
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr  
Roux, Paris 75015, France

## COMMENT

Collins clones from an A. gambiae BAC library provided by F. H. Collins and sequenced by Genoscreen in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.

## FEATURES

Location/Qualifiers  
1. .584

```
/organism="Anopheles gambiae"
```

```
/db_xref="taxon:7165"
```

```
/clone="32L10"
```

```

/clone_11b="Notredame1"
/note="end : T7"

```

a 180 c 153 g 1

BASE COUNT ORIGIN	107 a	180 c	153 g	138 t	6 others
----------------------	-------	-------	-------	-------	----------

## ORIGIN

Query Match	4.28;	Score 120.8;	DB 17;	Length 584;
Best Local Similarity	52.88;	Pred. No. 2.3e-11;		
Matches 274; Conservative	3;	Mismatches 241;	Indels 1;	Gaps 1;

OY	740	ATTGAAAGAGGATCGGGTGGCCCACTGGTACTGTTGGTTGTGTCTATTATTCATGCGAATG	799
	111		
	60	ACTGCGGCGAGTCCGGCTCCCGGACTGGAAAGCTGAGCCYCTGGCTCTCTATTCACG	119
OY	800	CTCATGGTGTTCGAGTGTGTCGCCCGAGAGTCAAGAGTTCCCGGCAAGCGGCCACTTC	859
	111		
Db	120	GTGGCGATGACCATCATGATGATGTGAAGGGATTCGGGATCCGGCAAGTGGCATCTTT	179
OY	860	CTCGCGCTCTCCCATACGTGTGATGATGATGATCTTTATTCATCACCAATATCTGCCC	919
	111		
Db	180	CTGGCCCTCCCCCGGACAGTGTGTCTGAACTCGATCGGCAATACGCTTACAGCTTGAG	239
OY	920	GGTCTACTGACGCGCATCTGTTCTTCTGTCACGCCCTCAATGGGGAACTCTTTGA-CTT	978
	111		
Db	240	GGCCCGGATCGGGGCTTGAAGTACTTATCATCACCCGACCGGATACAGCTCTGAATCG	299
OY	979	CGGGTATGGTACACAGTACAGACAGTCAAGATTTCTTCTCTGACAGTGTACCCGAC	1038
	111		
Db	300	GACGCTTGGAAGGAGGCCGTCTCGAGTGTCTTCTTCTGCTGTGATCTTTTTCGGGG	359
OY	1039	GATCATCATGTTCTCTCTTACACAGGTTTCAGACATAAATATTCACAGGAGTGGCTGGAT	1098
	111		
Db	360	TGTCATGCTTCTCTCTTACACAACTTCAGCAACAATATTACCGGGATGGAGAT	419
OY	1099	TGTTAGCACTTTGGACACCTTTACAAAGTTTCTTGTGGGTGACAGATCTTGGATCT	1158
	111		
Db	420	CATCTCGTGGCTGGATACGTTCACTCTCCCACTCTCCGGGGCGCTCGTGTTCATCAT	479
OY	1159	TGTTAACCTCGCGTACACATCACTACAGAGTGGAGATGTGGTGGTGTGGCGTAC	1218
	111		
Db	480	TGGCCTACTGGGACATCTGACAAACGAACGACATACCAAGTGGTAAAGCCGGGCG	539
OY	1219	CAGCTGTGTTTATTCATATACCTGATGCCATTGGCAA	1257
	111		
Db	540	TGGCTTAACTTATCACTACCTCCCGATGGGTGGGCCAA	578

Search completed: July 26, 2003, 03:07:04  
Job time : 3907 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2003, 23:43:13 ; Search time 131 Seconds

(without alignments)  
6695.386 Million cell updates/sec

Title: US-09-991-458-1

Perfect score: 2860

Sequence: 1 ggcacgaggtactactgtgtg.....ataaaaaaaaaaaaaaaaaa 2860

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	222	7.8	2403	2	US-08-700-013B-26
2	222	7.8	2817	1	US-08-655-836-4
3	222	7.8	2817	2	US-09-020-753-4
4	209.6	7.3	2728	1	US-07-879-617A-7
5	209.6	7.3	2728	1	US-08-753-985-7
6	209.4	7.3	3404	1	US-07-762-132A-1
7	204.8	7.2	3919	1	US-08-301-722A-1
8	196.4	6.9	2394	4	US-09-191-468-121
9	196.2	6.9	2278	1	US-07-959-943-6
10	196.2	6.9	2415	1	US-07-959-943-8
11	194.8	6.8	2394	4	US-09-191-468-119
12	193.2	6.8	2394	4	US-09-191-468-123
13	193.2	6.8	2397	2	US-08-700-013B-18
14	193.2	6.8	2397	2	US-08-700-013B-20
15	191.6	6.7	2863	4	US-09-182-718A-1
16	191.6	6.7	2863	4	US-09-795-232-1
17	189.4	6.6	2756	1	US-07-782-298-1
18	187.2	6.5	2508	1	US-07-959-943-10
19	161.4	5.6	2064	3	US-08-834-467-1
20	161.4	5.6	2064	4	US-09-336-177-1
21	161.4	5.6	2136	1	US-08-655-836-1
22	161.4	5.6	2136	2	US-09-020-753-1
23	161.4	5.6	2169	3	US-08-834-467-3
24	161.4	5.6	2169	4	US-09-396-177-3
25	161.4	5.6	2202	1	US-08-655-836-2
26	161.4	5.6	2202	2	US-09-020-753-2
27	161.4	5.6	2364	1	US-08-655-836-3

28	161.4	5.6	2364	2	US-09-020-753-3	Sequence 3, Appl1
29	160	5.6	1991	4	US-08-295-814E-9	Sequence 9, Appl1
30	160	5.6	1991	1	US-09-343-361-9	Sequence 9, Appl1
31	160	5.6	1991	5	PCT-US93-01959-9	Sequence 9, Appl1
32	156.2	5.5	1938	1	US-08-295-814E-3	Sequence 3, Appl1
33	156.2	5.5	1938	4	US-09-343-361-3	Sequence 3, Appl1
34	156.2	5.5	1938	5	PCT-US93-01959-3	Sequence 3, Appl1
35	154.8	5.4	2217	1	US-08-543-881-1	Sequence 1, Appl1
36	154.8	5.4	2217	1	US-08-291-299-1	Sequence 1, Appl1
37	154.8	5.4	2217	5	PCT-US94-00119-1	Sequence 1, Appl1
38	154.8	5.4	2217	5	PCT-US95-10579-1	Sequence 1, Appl1
39	151.6	5.3	2121	1	US-08-240-783B-1	Sequence 1, Appl1
40	151.6	5.3	2121	3	US-09-084-813-1	Sequence 1, Appl1
41	151.6	5.3	2121	5	PCT-US97-09662-1	Sequence 1, Appl1
42	148.4	5.2	755	4	US-09-191-468-71	Sequence 71, Appl1
43	147.6	5.2	2028	1	US-08-295-814E-1	Sequence 1, Appl1
44	147.6	5.2	2028	4	US-09-343-361-1	Sequence 1, Appl1
45	147.6	5.2	2028	5	PCT-US93-01959-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1

US-08-700-013B-26

Sequence 26, Application US/08700013B

Patent No. 5919653

GENERAL INFORMATION:

APPLICANT: Albert, Vivian R.

APPLICANT: Kowalski, Leslie R.Z.

APPLICANT: Borden, Laurence A.

APPLICANT: McKelvey, Jeffrey F.

TITLE OF INVENTION: Human Glycine Transporter

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTED for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,013B

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 317743-108

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 2403 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-700-013B-26

Query Match: 7.8%; Score 222; DB 2; Length 2403;  
Best Local Similarity 54.1%; Pred. No. 1.1e-44;  
Matches 503; Conservative 0; Mismatches 415; Indels 12; Gaps 2;

QY 746 GGAGTCTCGGTCGCCCATCTGTACTTGTGTGTATTCATCGATGGCTATG 805  
 Db 1165 GAATATCTGCTGAGATCAGATGGCCCTTCCGCTTTTCCGCGCTGGGATAT 1224  
 QY 806 GTGTGGAGTCTGCGCCGAGATCAGAGTTCGGCAAGCGGCTACTTCTCGG 865  
 Db 1225 GTATATGCAATCGCTGGCAAAAGATTAAGCATCAGAAAGATGGTACTTACAGCC 1284  
 QY 866 CTCTCCCATAGCTGTGATGATCAGTTATTCATACCAACATCCGCGGCTGT 925  
 Db 1285 ACCCTCCCTTATGTCTGCTGCTGATCTCTCATTCGAGGGCTACCTGCTGGAGCT 1344  
 QY 926 ACTGAGGACCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985  
 Db 1345 GGAGCCGGTATCTGATCTATCATCAGACCTTAAGTGGGAGAACTCAGCGATGCGAGCTG 1404  
 QY 986 TGTACTCAGCATCAGCAGAGTGTCTTCTCTGACAGTGTGACCGGACCGATCATC 1045  
 Db 1405 TGGAGGATGCAACCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1464  
 QY 1046 ATGTCTCTCTTACAAAGCTTTCAGACATATATATACAGGGATGCTGATGTTAG 1105  
 Db 1465 ACTCTTCTCTTACAAAGCTTTCAGACATATATATACAGGGATGCTGATGTTAG 1105  
 QY 1106 ACTTGGACACCTTACAAAGCTTTCAGACATATATATACAGGGATGCTGATGTTAG 1165  
 Db 1525 TGCACCAACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1584  
 QY 1166 CTGCGGTACGACATCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1225  
 Db 1585 ATGGCCACGAGGCGAAAGTCAACATTTGAGAAATGTG---GCTGACCAAGGCGGAGGAT 1641  
 QY 1226 GCTTTCATTCATACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1285  
 Db 1642 GCATTTGTGTTTACCAAGAGCTTACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1701  
 QY 1286 CTGCTCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1345  
 Db 1702 ATCTTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1761  
 QY 1346 TTCAACATTTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1405  
 Db 1762 ATGTGACCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1821  
 QY 1406 AC-----CTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1456  
 Db 1822 ACCCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1881  
 QY 1457 CAATATATCTTGAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1516  
 Db 1882 ATCTACATGTTTGAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1941  
 QY 1517 ATTTGGAACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1576  
 Db 1942 ATTTTGAAGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2001  
 QY 1577 GAGTATGTTGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1636  
 Db 2002 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2061  
 QY 1637 CCTGCTATATGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1666  
 Db 2062 CCGACCATTTTAAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2091

## RESULT 2

; Sequence 4, Application US/08655836  
 ; Patent No. 5824486  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Borden, Laurence A.  
 ; APPLICANT: De Vivo, Michael  
 ; APPLICANT: Yokoyama, Midori

; APPLICANT: Albert, Vivian R.  
 ; TITLE OF INVENTION: GLYCINE TRANSPORTER-TRANSFECTED CELLS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dechert Price & Rhoads  
 ; STREET: P.O. Box 5218  
 ; CITY: Princeton  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 08543  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patenting Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/655,836  
 ; FILING DATE: 31-MAY-1996  
 ; CLASSIFICATION: 436  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bloom, Allen  
 ; REGISTRATION NUMBER: 29135  
 ; REFERENCE/DOCKET NUMBER: 317743-105  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 609-520-3214  
 ; TELEPHONE: 609-520-3259  
 ; INFORMATION FOR SEQ. ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2817 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-08-655-836-4

Query Match 7.8%; Score 222; DB 1; Length 2817;  
 Best Local Similarity 54.1%; Pred. No. 1.2e-44;  
 Matches 503; Conservative 0; Mismatches 415; Indels 12; Gaps 2;

QY 746 GGAGTCTCGGTCGCCCATCTGTACTTGTGTGTATTCATCGATGGCTATG 805  
 Db 1372 GAATATCTGCTGAGATCAGATGGCCCTTCCGCTTTTCCGCGCTGGGATAT 1431  
 QY 806 GTGTGGAGTCTGCGCCGAGATCAGAGTTCGGCAAGCGGCTACTTCTCGG 865  
 Db 1432 GTATATGCAATCGCTGGCAAAAGATTAAGCATCAGAAAGATGGTACTTACAGCC 1491  
 QY 866 CTCTCCCATAGCTGTGATGATCAGTTATTCATCAGCAACATTCCTGCGGCTGT 925  
 Db 1492 ACCCTCCCTTATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1551  
 QY 926 ACTGAGGGAATCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985  
 Db 1552 GGAGCCGGTATCTGATCTTCAATCAGACCTTAAGTGGGAGAACTCAGGCTGCGAGCTG 1611  
 QY 986 TGTACTCAGCAGTCAAGAGTGTCTTCTCTGACAGTGTGACCGGACCGATCATC 1045  
 Db 1612 TGGAGAGATGACACCACTGATTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1671  
 QY 1046 ATGTCTCTCTTACAAAGGTTTCAAGATATATATATATATATATATATATATATAT 1105  
 Db 1672 ACTCTTCTCTTACAAAGGTTTCAAGATATATATATATATATATATATATATATAT 11731  
 QY 1106 ACTTGGACACCTTTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1165  
 Db 1732 TGCACCAACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1791  
 QY 1166 CTGCGGTACGACATCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1225  
 Db 1792 ATGGCCACGAGGCGAAAGTCAACATTTGAGAAATGTG---GCTGACCAAGGCGGAGGAT 1848  
 QY 1226 GCTTTCATTCATACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1285



RESULT 4  
US-07-879-617A-7  
Sequence 7, Application US/07879617A  
Patent No. 5580775  
GENERAL INFORMATION:  
APPLICANT: Fremieu Jr., Robert T.  
APPLICANT: Caron, Marc G.  
APPLICANT: Blakely, Randy D.  
TITLE OF INVENTION: A High Affinity L-Proline Transporter  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/879,617A  
FILING DATE: 19920501  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2728 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus  
TISSUE TYPE: Brain  
IMMEDIATE SOURCE:  
LIBRARY: rat forebrain cDNA library  
CLONE: rTB2-2-20  
US-07-879-617A-7

Query Match 7.3%; Score 209.6; DB 1; Length 2728;  
Best Local Similarity 48.2%; Pred. No. 1.2e-41;  
Matches 707; Conservative 0; Mismatches 729; Indels 30; Gaps 3;

QY 260 TGGTCCACACATGTAATTCCTGATGTCCTGCATCATACCGTGGTTGGTAA 319  
DB 208 TGGACGGGCAAGCTGACCTCTGCTGCTCATGCGGCTACGTGAGGCTTGGGAAT 267  
QY 320 GTGTGGCGGTTCCTTCATCGGGTACGAAATGAGGAGAGGAGGCTTCCGCGGCATAC 379  
DB 268 GTCTGGCGGTTTCCTATGAGGCTTACACCAATGAGGAGGAGGAGGCTTCCGCGGCATAC 327  
QY 380 GTGATCGTCTTACTTGTGCGCAAGCGTGTACTACTTACTAGTGTGCTCGACAA 439  
DB 328 TTCCCTCATGCTGGCCATCTGTGSCATCCGCCCTTCTTTCTTACGCTCTCCGCGGCAG 387  
QY 440 TTGAGTTCAAGAACTCTGTTAAAGTTGGTCAATTTACCGGCGCATGAAGGTA 499  
DB 388 TTCTCCAGCGCTGGAGACCCCTGGCGTGTGGAATAACAGCCCTCTTCAAGGTCGCGGT 447  
QY 500 TACGCTCAAGCTGCGCGCTGCGGTACATCCCTGCTTACTACGTGGTGAATCTGGTCTC 559

DB 448 GCAAGCATGCTGCTCATCTGGGCGCTGGTGGCCATCTACTACAAACATGATCATCGCCATC 507  
QY 560 TGTCTATTACTTACTAGTATGAGCTTCCAGGCCACTTCTCCATGCGCTATTGTACGCT 619  
DB 508 GTCTCTTCTTACCTCTTGGCTCCCTCCACAGAACCTGCGGGAGCATTTGGGCGCAC 567  
QY 620 GAGTG-----GGAGACGTGGGTACCTCCAGATCCCAACACTTGTGCA 661  
DB 568 TGTGTGAACACAGAGCGCTCCGAGACACAGAGGCCCAAGATGGCAACGGGCGACTG 627  
QY 662 TCAGTCAACACATCACCAGATGATCCAGAGTCTCACTCACTTTTGAACAGTT 721  
DB 628 CTTCTTACCTCAGACACTGTCAGGCCAGTAGAGAGTACGAGCCGATATGTCTG 687  
QY 722 CTCACCAAGCGATGGAATGAAGAGTCTGGTCCGCCCATCTGTGATGTTG 781  
DB 688 CACATTCAGGGCAGCCAGGCGATCGGCCAGCCGGGAGATTCGCTGGAACCTTGCTC 747  
QY 782 TGTCTATTATGCGATGCTCATGATGTTGGAGTGTGCGCCGAGGAGTCAAGATTC 841  
DB 748 TGCCTGTGCTGGCTGGCTGGCTGATCTGTTCTGTTCTGTTCTGTAAGAGGAGGATGCTC 807  
QY 842 GGCAGAGCGGCTTACTCTCGGCTCTTCCATACGTTGCTCATGATCATTTATTCATC 901  
DB 808 GCGAAGGTGGTATTTACAGCACCTTCCCTACCTCATCTGCTCATGCTGCTGCT 867  
QY 902 ACCACATCATCTCCGCGCGGCTGCTACAGAGGATCTGTTCTTCTGACGCTCATG 961  
DB 868 CGAGAGTGAACCTTCCCGGGGCTGGAGAGGATCCAGTTTATCTACCCGCCCATTC 927  
QY 962 GCGAATCTCTGAGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1021  
DB 928 CACCACTGCTGCTTCTTCCAGAGTGTGATGATGATGATGATGATGATGATGATGAT 987  
QY 1022 ACAGTGTGACCGGACGATATGTTCTCTCTTACAGGTTTACAGATTAATATC 1081  
DB 988 GGAATGGGTTTGGGGCTTCTTCCACCTTCTTCCACCAACATTCACCAAGATC 1047  
QY 1082 TACAGGAGCTTGGATGTTTACAGATTTGAGACCTTTCAGATTTCTGTGGGTG 1141  
DB 1048 TACAGAGACACCTTATGCTATGCTATGCTGAGGATGAGGATGAGGATGAGGATG 1107  
QY 1142 AGATCTTGGTATCTTGTGTAACCTGCGTACGAACTCACTCAGAGTGGAGATG 1201  
DB 1108 GCTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1167  
QY 1202 GTGCGTCTGCGGTACGATCTTGTGTTTATTTATACCGCTGATGCTGCAAAACA 1261  
DB 1168 GCGAAGCAGGC---CCTGGCTGGCTTGTATCTACCCAGGCGATGATGTTG 1224  
QY 1262 TTCCAACTGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1321  
DB 1225 CCTCTGACACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1284  
QY 1322 TCATCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1381  
DB 1285 AGCCAGTTGCTTCTGGAACATGATGATGATGATGATGATGATGATGATGATGATG 1344  
QY 1382 CCGACCGCTACATGTCAGCA-----TACCGTCTGCTGCTGCTGCTGCTGCTGCTG 1432  
DB 1345 CTAGGCGCCCAAGAGGAGTGTCTGAGGCTCATCTGCTGATGCTGATGATGATGATG 1404  
QY 1433 CTGTTTACTGACACCGGCTGAGCAATATATCTGAGCTTATATCATCTAGCTGGA 1492  
DB 1405 CTGATCTCAGACACCGATGAGGAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTG 1464  
QY 1493 ACATCTCTGCTTCTTCTGCGCATTTCTGAACCTGCGAGCGTGTCTGATTTATGA 1552  
DB 1465 AGCTTGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1524  
QY 1553 TTGGAGATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1612







QY 1064 GATTTCAGACATTAATATCTAACAGGATGCTTGAGATTGTTAGACCTTTGAGACCTTTACA 1123  
 Db 1068 AATATTCACAAATTAACCTGCTATATAGAGAGCAATCAATCAACACCTTCATTAATCTCCAGCA 1127  
 QY 1124 AGTTTCCTTGCTGGGTGACAGANTCTTGAGATTCCTTGATTAACCTGAGCTAGCACTAAC 1183  
 Db 1128 AGCTTCCTGCTGCTTCGTGCTGCTTCCTTCCTTCCTGGGGATATATGAGCAAGAACACAT 1187  
 QY 1184 TCAGAGGTGGAGATATGCTGCGGTGCTGGGGGTACAGCTCTTGCTTTCAATTCATACCT 1243  
 Db 1188 GTGCCCATCAAGAGATGTGGCC--ACAGATGGACCTGAGGCTCAATTCATCATATACCT 1244  
 QY 1244 GATGCCAATTTGCAAAAATTTCCAACTCAACCTCAGACTATTCGCGGTGCTTCCTCCATGAT 1303  
 Db 1245 GAGCGAGATGCGACACATCCCGCTGTCTTTGCTCCGGGCTGTGTCTTCCTTCCTATGCTG 1304  
 QY 1304 TCGGTGCTGGGTATAGCGCTCATCCGTGGCTCTGCTATGCACTTTCAACACATTTGGCGATG 1363  
 Db 1305 CTCACCTCTGGGTATCGACAGTGCAMTGGGGGGCATGAGATCAGTATCACTGGGGCTCGTC 1364  
 QY 1364 GAGCGTTCACAGCTGATAC--CCACCGTCTACATGACAGATGACCTGCTTCGCG- 1418  
 Db 1365 GATGATGTTCCAGCTGCTATACATCGGCATCGAGAGCTCTTCACTCTTGCGCATTTGCTGGCT 1424  
 QY 1419 -GTTTCCTGCTTGACCTGTTTACTGACACACCGGGTGAGACAAATATATCTTGAGACTTGA 1477  
 Db 1425 ACTTTCCTGCTGCTCTCTCTCTGCTGCTGACACAGGTGGACATCAGCTTCACACTGCTG 1484  
 QY 1478 GATACATACGCTGGAACATTCCTGCTGCTTTCTTGCGGCATTTCTGAACTGCGAGCGCTG 1537  
 Db 1485 GACCATCTTCAGCTGCGACATCATATCTTTTGGCGCTGCTCATTGAAAGCATTTGGGGCTG 1544  
 QY 1538 TTCTGATTTATGATTTGAGAAATCTGTGCTAGACATTTGATTCATGTTGGGTGAAAG 1597  
 Db 1545 GCGTGGTTTACGGCGTCCAGCAATTCAGATGATGACATCAAGCAAAATGACAGGGCAGCA 1604  
 QY 1598 ACTGCTGCTTACTGCGCTCTCTGCTGGGGGCTAATCACTCTCTGATATATGACAGACTGTG 1657  
 Db 1605 CCCAACCTGATGCTGGCGGCTAATACTGGAAGCTGTGACGCCCTGCTTCCTCTGATGTG 1664  
 QY 1658 TTCTTACAGCTCTTTCGCGCTCTCAACACCTGTGTTGCGAGACACTGCT 1710  
 Db 1665 GTGCTGTTACGANTGTGACCTTCACAGCCCCACACTATATGGGGCTCATATCTT 1717

RESULT 7  
 US-08-301-722A-1  
 : Sequence 1, Application US/08301722A  
 : Patent No. 5756307  
 GENERAL INFORMATION:  
 APPLICANT: UHL, George R.  
 APPLICANT: Vandenberg, David  
 APPLICANT: Persico, Antonio  
 TITLE OF INVENTION: SEQUENCE OF HUMAN DOPAMINE TRANSPORTER  
 TITLE OF INVENTION: CDNA  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
 STREET: P.O. Box 747  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22040-3487  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/301,722A  
 FILING DATE: 07-SEP-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:

[illegible]

OY	821	GCCGAGGAGACGACAGAGTTCGGGCAAGCGGCGCTACTTCCCTCGGCGCTCCGCCATACGTT	880
Db	867	TGGAAGGCGCGGAAGACCTCAGGGGAAGGTGATAGATACAGCACCACATCATTACGTG	926
OY	881	GTCATGATCATTATTCATCACCACATCATCTCCGCGGTGCTACTGACGCGATCCTG	940
Db	927	GTCCTCATGCGCCCTGCTCCTCGCTGGGGATCACCCTCCCTTGAGGCATATGACGCATCGA	986
OY	941	TTCTTCGTCACGCGCTCAATGGGCGCAACTCCTTGAGCTCGGTGATAGTACTCAGCACTC	1000
Db	987	GCATACCTCGACGCTTGACTTCTTACCCGGCTCTCGCAGGCGCTGTGGTTGATTAGCGCGCC	1046
OY	1001	ACGCAAGGTCTTCTCTCTCTGACAGTGTGCACCGGACGATCATCATCTTCTCCTCTAC	1060
Db	1047	ACCCAGGTGCTCTTCTCCTCGCTGGGCGTGGGGATTCGGGGTCTGATGCCCTTCTCCAGCTAC	1106
OY	1061	AACGGTTTCAGACATATATCTACAGGAGTCTTGAGTTGTTACGACTTGGACACTTT	1120
Db	1107	AACCAAGTTCACCAACACACTGCTTACGAGGACCGCATTTGTCACACCTCCATCACTCCCTG	1166
OY	1121	ACAAAGTTCTGTGTGGGTGACAGCATCTTGCGTATCTTGCTTAACTCCGCGTACGAATCT	1180
Db	1167	ACGAGCTTCTCTCTCGGCTTGTGTGCTCTTCTCTTCTCTCGGGGTACATGGCACAGAACAC	1226
OY	1181	AACCTCAGAGTGGGAGATGTGTGCGTCTGCGGTACACAGTCTTGTCTTCAATTTCAATAC	1240
Db	1227	AGTGGCCCATCGGGGAGCTGGGC---AAGACGCGGCCACGAGGCTGATCTTCATCATCTAC	1283
OY	1241	CCTGATGCGCATTTGCCAAACAACTTCCACACTCAGCTATCTCGGGTGTCTTCTCTGATG	1300
Db	1284	CCGGAAAGCCATACGCCACGCTCCCTGTGTCTTCACGCTGGGCGCGTGGTCTTTCATCATCTG	1343
OY	1301	ATGTGCGTGTGGGTATCGGCTCATCTCCGTGCTGTCTGTATCGACTTTTCACACATTTGGCG	1360
Db	1344	CTGCTCACCTCGGGGATGACAGCGCCCATGGGTGGTATGAGTACGATGATCAGCGGGCTC	1403
OY	1361	ATGGAAGCGTTCCAC----GTGTAACCAACCGTCTACATGTGCAGCGATGACCTGTTCTTG	1416
Db	1404	ATCGATGAGTTTCACACTCTGTCACAGACCGGTGAGCTCTTACGCGCTTATCGTCTCG	1463
OY	1417	CGGT--TTTCGCTGTGAGCTATTGTACTGACACACCGGGTGGAGCAATATATCTTGAGCTT	1474
Db	1464	GCGACCTTCTCTCTGTCTCCCTGTGTCTGCTGTACACAAAGGTGGCATCTACGTTCTACGCTC	1523
OY	1475	GTAGATCACTACGGTGAACATCTCTTGTTGCTTTTGTGCGCAATTTGAACTCGCAGGC	1534
Db	1524	CTGGACCACTTTTGGACCGCGCACGTCCATCTCTTTGGAGTGCATATGGAAGCATCGGA	1583
OY	1535	GTTTCTCGATTTTATGATTTGGAGAACTGTGGCTAGACATTTGATGATGTTGGGTAA	1594
Db	1584	GTGGCTGTCTTATATGTGTGTGGCAGTTACACGACGACATTCACAGATACCGGGCAG	1643
OY	1595	AAGACTGTGCTTACTGTGCGCTGTCTGTGGGGCGTAACTCCTGCTATATAGACACT	1654
Db	1644	CGGCCCACTCTACTGTGCGGCTGTGCTGGGAAGTGGCACCCCTGCTTCTCCTGTTTC	1703
OY	1655	GTTGTTCTCTTACGCTCTTCTGCGCCCTTATCAACAACCTGGTGTTCGGAGAACACTAGCT	1710
Db	1704	GTGTCGTGTACGATTTGTACCTTCAGACCCCCCCTACTACGGAGGCTTACATCTT	1759
RESULT 8			
US-09-191-468-121			
; Sequence 121 Application US/09191468A			
; Patent No. 641695			
; GENERAL INFORMATION:			
; APPLICANT: Gallagher, Michael J.			
; APPLICANT: Burgess, Loyd R.			
; APPLICANT: Brunden, Kurt R.			
; TITLE OF INVENTION: Human Glycine Transporter Type 2			
; FILE REFERENCE: 1231US01			
; CURRENT APPLICATION NUMBER: US/09/191,468A			
; CURRENT FILING DATE: 1998-11-12			

[illegible]









```

Db 1219 GTGTATGATCGTGGCTAAAGATACAGACTTCAGAAAAAGTGTACTTACAGGCC 1278
QY 866 CTCTTCCCATACAGCTTGTATCATCTTATTCATCCACAAATCATCTCCCGGGTCT 925
Db 1279 ACCTTCCCGATATGTCTGATCTGATCTCTTCTTCCGAGAGATCCCTGCGGAGCT 1338
QY 926 ACTGAGGACATCTGTTCTTCTGTCAGCCCTCAATGGCGAAGTCTCTTGAAGTGTGA 985
Db 1339 GGAGCTGGATCTGTTCTTCTTATCACCACCAAGTGGAGAAAGTCAAGAGTGCACAGCTG 1398
QY 986 TGTATCAGACATCAGCAGAGTGTCTCTCTGACAGTGTGACCCGAGCCAGATCATC 1045
Db 1399 TGAAGAGTCTCCCATCTGATTTCTTCTTATCTCTGATGAGGAGGCGCTGATC 1458
QY 1046 ATGTCTCTCTTACAAAGGTTTCTGACATATATCTACAGAGGATGCTTGAATTTGACG 1105
Db 1459 ACTCTCTCTTCTTACAAATTTCCACAACTGCTACAGGAGACTCTAATTTGTACCC 1518
QY 1106 ACTTTGGACACCTTTACAGATTTCTGTCTGGGTGACAGATCTTGGTATCTTGTATAC 1165
Db 1519 TGCACCAACAGTGCACCAACATCTTGGCGCTTGTCTGATCTCTGATTCGAGCTTC 1578
QY 1166 CTGGCGTACAGCACTCACTAGAGTGGAGATGTGTGCTGCTGCGGTACCACTT 1225
Db 1579 ATGGCCATGATGCAAGAAATCAATTTGAGATGTG--GCAAGACCAAGGGCCAGGCTT 1635
QY 1226 GCTTTCATTTTCAATCCCTGATGCTATGCTCAAAATTTCCACCTGATCTTCTGCTG 1285
Db 1636 GCATTTGTGTGTTATCCCGAAGCTTTAACCAAGCTGCTCTCTCTCTGCTTCTGCGCATC 1695
QY 1286 CTGTCTCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1345
Db 1696 ATCTTTTCTCTGATGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1755
QY 1346 TTCAACGATTTGGAGATGAGCGGGTCCACAGTGTACCCAGCTGATCATGATGATGATG 1405
Db 1756 ATGTGATCTCTCTGATGAGCGAGTTCCTTCCCAAGTACTTACGACACACCAAGCTGTTT 1815
QY 1406 AC-----CTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1456
Db 1816 ACTGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1875
QY 1457 CAATATATTTCTTGAAGTGTGATGATCAGTGGGAGCAATTCCTTGTGCTTCTGCGCC 1516
Db 1876 ATTTAATGTTTGAAGTGTGATGATCAGTGGGAGCAATTCCTTGTGCTTCTGCGCC 1935
QY 1517 ATTTTGAATCTGACGAGCGGTCTGATTTATGATTTGAGATGAGATGCTGCTGATGAT 1576
Db 1936 ATTTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1995
QY 1577 GAGTTCATGTTGGGTAAAGAAAGTGTGCTTACTGCGCTCTGCTGCGGCGGTATCATC 1636
Db 1996 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2055
QY 1637 CCGTATATATGACGACTGTGTCTTCTTAC 1666
Db 2056 CCAACCAATTTTAACTTTATCTTGTCTTC 2085

```

```

RESULT 13
US-08-700-013B-18
; Sequence 18, Application US/08700013B
; Patent No. 5919653
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian R.
; APPLICANT: Kowalski, Leslie R.2.
; APPLICANT: Borden, Laurence A.
; APPLICANT: McKeiv, Jeffrey F.
; TITLE OF INVENTION: Human Glycine Transporter
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210

```

```

; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,013B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2397 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-700-013B-18

```

Query Match 6.8%; Score 193.2; DB 2; Length 2397;

Best Local Similarity 52.2%; Pred. No 1.2e-37;

Matches 485; Conservative 0; Mismatches 433; Indels 12; Gaps 2;

```

QY 746 GGAAGTCTCGTGCCCGCCATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 805
Db 1159 GAATATCTGTCAGATCAGATGAGTGGCCACTGCTCTGCTCTGCTCTGCTGCTGCTGCT 1218
QY 806 GTGTGCGATCTGCTGCGCGAGAGTCAAGATTCGGCAAGCGGCTCTCTCTGCTGCTG 865
Db 1219 GTGTGATGATCTGTTGGTAAAGATCAAGATTCAGGAAAGTGGTGTCTTCTACAGGCC 1278
QY 866 CTCTTCCCATACAGCTTGTATGATCACTTTATTCATCCCAATTCATCTGCGCGTGTCT 925
Db 1279 ACCTTCCCGATATGTCTGATCTGATCTCTCTCTATCTCTGATGAGAGTACCTGCGGAGCT 1338
QY 926 ACTGAGGACATCTCTGTTCTTCTGTCAGCGCTCAATGGCGAAGTCTTGTGAGTGTGTA 985
Db 1339 GGAGCTGGGATCTGTTCTTCTGATCAACCCAGTGGGAGAAAGTCAAGGATGCGACGCTG 1398
QY 986 TGTATCAGACATCAGCAGAGTGTCTCTCTGACAGTGTGACCCGAGCCAGATCATC 1045
Db 1399 TGAAGAGTCTGCTGCTGCTGATTTCTTCTTATCTCTGATGAGGAGGCGCTGATC 1458
QY 1046 ATGTCTCTCTTACAAAGGTTTCTGACATATATCTACAGAGGATGCTTGAATTTGACG 1105
Db 1459 ACTCTCTCTTCTTACAAAGTTCACAACTGCTACAGGAGACTCTAATTTGTACCC 1518
QY 1106 ACTTTGGACACCTTTACAGATTTCTGTCTGGGTGACAGATCTTGGTATCTTGTATAC 1165
Db 1519 TGCACCAACAGTGCACCAACATCTTGGCGCTTGTCTGATCTCTGATTCGAGCTTC 1578
QY 1166 CTGGCGTACAGCACTCACTAGAGTGGAGATGTGTGCTGCTGCGGTACCACTT 1225
Db 1579 ATGGCCATGATGCAAGAAATCAATTTGAGATGTG--GCAAGACCAAGGGCCAGGCTT 1635
QY 1226 GCTTTCATTTTCAATCCCTGATGCTATGCTCAAAATTTCCAACTGATCTGATCTGCTG 1285
Db 1636 GCATTTGTGTGTTATCCCGAAGCTTTAACCAAGCTGCTCTCTCTGCTTCTGCGCATC 1695
QY 1286 CTGTCTCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1345

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 02:02:13 ; Search time 599 seconds  
(without alignments)  
9850.054 Million cell updates/sec

Title:	US-09-991-458-1
Perfect score:	2860
Sequence:	1 ggcacgaggttactgttg.....ataaaaaaaaaaaaaa 2860

Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	204.8	7.2	3946	15	US-10-213-948-4	Sequence 4, Apple
2	191.6	6.7	2863	10	US-09-795-232-1	Sequence 1, Apple
3	190.4	6.7	1764	11	US-09-815-923-3	Sequence 3, Apple
4	187.2	6.5	4305	15	US-10-198-846-11010	Sequence 11010, Apple
5	176.4	6.2	4956	11	US-09-815-923-9	Sequence 9, Apple
6	162	5.7	3513	11	US-09-815-923-13	Sequence 13, Apple
7	160.6	5.6	2283	11	US-09-954-456-160	Sequence 560, Apple
8	160.6	5.6	2283	11	US-09-954-456-1630	Sequence 1630, Apple
9	154.8	5.4	2219	12	US-09-919-039-377	Sequence 377, Apple
10	154.8	5.4	3410	11	US-09-880-107-3339	Sequence 3339, Apple
11	147.6	5.2	2028	11	US-09-917-800A-1429	Sequence 1429, Apple
12	146.8	5.1	2016	10	US-09-843-598-1	Sequence 1, Apple
13	146.8	5.1	2016	10	US-09-843-598-3	Sequence 3, Apple
14	145.2	5.1	2016	10	US-09-843-598-2	Sequence 2, Apple
15	126.6	4.4	1671	10	US-09-815-923-2	Sequence 5, Apple
16	122.6	4.3	1798	11	US-09-818-656A-1	Sequence 1, Apple

15	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17
55	55	55	54	56	56	57	57	58	58	59	61	65	65	86.2	87.2	96.4	98	100.4	105	105	105	105	105	105	105	105	102	122.6
1.9	1.9	1.9	1.9	2.0	2.0	2.0	2.0	2.0	2.1	2.1	2.1	2.3	2.3	3.0	3.0	3.4	3.4	3.5	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.8	4.3	
3252	3252	3252	424	424	5689	5935	1830	755	1431	1714	5979	472	628	2657	451	758	2034	2485	1833	3103	3103	3103	3103	2190	2190	2190	1798	
US-10-027-632-113788	US-10-027-632-113788	US-10-027-632-113788	US-09-960-352-11218	US-10-239-676-1210	US-10-239-676-134	US-10-329-960-1	US-10-027-632-128109	US-09-774-414-2	US-10-239-676-206	US-10-339-676-18	US-09-918-993-34	US-10-027-632-113152	US-10-313-542-191	US-09-918-995-5064	US-09-910-943-629	US-09-815-923-11	US-09-923-444A-1	US-09-861-846-1	US-10-156-239-4	US-10-156-239-4	US-10-156-239-4	US-10-156-239-6	US-10-156-239-6	US-10-156-239-6	US-10-156-239-6	US-09-843-598-4	US-10-216-441-1	
Sequence 113788,	Sequence 113787,	Sequence 113786,	Sequence 11218, A	Sequence 210, App	Sequence 90, App	Sequence 134, App	Sequence 128109,	Sequence 2, App	Sequence 106, App	Sequence 18, App	Sequence 34, App	Sequence 113152,	Sequence 191, App	Sequence 5064, App	Sequence 629, App	Sequence 11, App	Sequence 11, App	Sequence 4, App	Sequence 4, App	Sequence 4, App	Sequence 4, App	Sequence 6, App	Sequence 6, App	Sequence 6, App	Sequence 6, App	Sequence 4, App	Sequence 1, App	

## ALIGNMENTS

```

RESULT 1
US-10-213-948-4
: Sequence 4, Application US/10213948
: Publication No. US20030100479A1
: GENERAL INFORMATION:
: APPLICANT: SmithKline Beecham
: TITLE OF INVENTION: Gene Polymorphisms and Response to Treatment
: FILE REFERENCE: pg4541
: CURRENT APPLICATION NUMBER: US/10/213,948
: CURRENT FILING DATE: 2002-08-07
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 3946
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-213-948-4

```

Query Match	7.2%	Score 204.8	DB 15	Length 3946
Best Local Similarity	48.6%	Pred. No. 1e-33		
Matches 727	Conservative	0	Mismatches 742	Indels 27
				Gaps 5

[illegible]

Db 534 CCCATCTGAAAGGTGGGCTTCACGGTCACTCCTATCTCACTGATGTCGGCTCTTC 593  
QY 539 TACGTGTGATCTGTGGTCTCTGTCTGATTAAGTATGATGCTCCAGGCACCTT 598  
Db 594 TAAACGTCTATCTCCGCTGGCGCTGCTACTATCTCTCTCTCTCTCTCTCTCT 653  
QY 599 CCAATGGGCTATTTGTGACCTGATGGAGAACCTGGTACCTCAGATCCAGACTTGT 658  
Db 654 CCTGTGATCCACTGCAACCACTCTGGAACAGCCCAACGTCGTGGATGCCATCTGGT 713  
QY 659 G-----CATGATCACAACATCATCCATGTTACCACTGCTCAATC 703  
Db 714 GACTCCAGTGGAGACAGCTGGGCTTCACAGACATTTTGGAGCACACCTGTCGCGAG 773  
QY 704 TACTTTTGGAGACAGTCTCCAAACAAGGATGSAATG---AAGGAGGCTCCGCGCC 760  
Db 774 TACTTTGAAAGTGGCGCTGCTGCACTCCACCAAGGCCATGCGATCGACGCTGGGCGCT 833  
QY 761 CCCATCTGTACTGTGTGTGTCTATTCATCGATGCTCATGATGTTCCGAGTGTG 820  
Db 834 CCGCGGTGGAGCTGACAGGCTGCTGGTGTCTGTATCTGCTGCTCTACTTCAAGCTC 893  
QY 821 GCGCGAGGATCAAGATTCGCGCAAGCGGCTACTTCTCGCGCTCTTCCATACGTT 880  
Db 894 TGGAGGGCGTGAAGACCTCAGGAAGGTGTATGATCAACACCAACATGCTCATAGCTG 953  
QY 881 GTCATGATCACTTATATCATCACCAACATCATCTGCGCGGCTCTACTGCGGATCTG 940  
Db 954 GTCTCTACCTGCTGCTCTGCTGGGCTACCTCTCTCTGAGCATGACGCGCATCAGA 1013  
QY 941 TTTCTGTACAGGCTCTCAATGGGCAACCTCTTGAAGCTGCTGTATGTTACTGACAGTC 1000  
Db 1014 GCATACCTGAGGCTTACTCTTCACTCTACCGGCTCTGCGAGGCGCTGTGGATGACGCGGCC 1073  
QY 1001 AGCAGATGTTCTTCTCTCTGACAGTGTGACCGGACCATCATGTTCTCTCTTAC 1060  
Db 1074 ACCGAGGTGTCTCTCTCTGCGGCTGGGCTGGGCTGTGATCGCTCTCTCAAGTAC 1133  
QY 1061 AACGTTTTCAGACATATATCTACAGGATGTTGATGTTGATGATGATGATGATGATG 1120  
Db 1134 AACAAATTCACCAACAACATCTCTACAGGACGATGTCACCACTCTCACTCACTCTG 1193  
QY 1121 ACAAGTTTCTGCTGTGGGTGACAGATCTTCTGATCTCTGTTACCTCGGTCAGATC 1180  
Db 1194 ACAGGCT 1253  
QY 1181 AACTCAGAGGTGGAGATGTTGTCGGTGTGGGCTGACCAAGTCTTCTCTCTCTCTCTCT 1240  
Db 1254 AGTGTGCCATTCGGGACGTGGCC---MAGGACGGGCGAGGCTGATCTTCACTCATCTAC 1310  
QY 1241 CCTGATGCAATTCGCAAAACATTCACACCTCAGCTATTTCTGCTGTTCTCTCTGAT 1300  
Db 1311 CCGGAAGCATTCGCAAGCT 1370  
QY 1301 ATGTGCTGTGGGTATCGGCTCATCGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1360  
Db 1371 CTGTCTACCTCTGTGTGACACAGCTGATGGGTATGATGATGATGATGATGATGATG 1430  
QY 1361 ATGAGCGGCTTCCAC---GTGTACCAACGCTCTCATGATGATGATGATGATGATGATG 1416  
Db 1431 ATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1490  
QY 1417 CCGT---TTCTGCTTGGATGTTTATCTGACACCGGCTGGACATATATTTCTGAGCTT 1474  
Db 1491 GGGACCT 1550  
QY 1475 GATGATCACTAGGCTGGAACATTTCTTGTCTTTTCTGCGCATTTTCTGAACTCGAAGC 1534  
Db 1551 CTGGACATTTTTCACCCGCAAGCTCATCTCTTGTGATGCTCTCATGAGCAATCGGA 1610  
QY 1535 GTGTTCTGATTTATGATTTGGAATCTGTGCTGATGATGATGATGATGATGATGATGAT 1594

Db 1611 GTGGCCTGATTTATGATGTTGTGGCAATTCAGCGACACATCCAGCATGATGACCGGCGAG 1670  
QY 1595 AAGAGTGTGCTTACTAGTGGCTCTCTGTGGGGGTATTCATCTCTCTATATAGAGACT 1654  
Db 1671 GGGCCAGCTGTACTGCGCGCTGTGCTGGAAGCTGTGATGACGCCCTGTCTCTCTCT 1730  
QY 1655 GTGTTCTTTCACGCT 1710  
Db 1731 GTGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1786

RESULT 2  
US-09-795-232-1  
; Sequence 1, Application US/09795232  
; Patent No. US20010012627A1  
; GENERAL INFORMATION:  
; APPLICANT: Anthony M. Brown  
; APPLICANT: Conrad Gerald Chapman  
; APPLICANT: Israel Simon Glover  
; APPLICANT: Joanne Rachel Evans  
; APPLICANT: William Cairns  
; APPLICANT: Hugh Jonathan Herdon  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30176-D1  
; CURRENT APPLICATION NUMBER: US/09/795,232  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 09/182,728  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 9818890.7  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2863  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-795-232-1

Query Match 6.7%; Score 191.6; DB 10; Length 2863;  
Best Local Similarity 52.0%; Pred. No. 6e-31;  
Matches 484; Conservative 0; Mismatches 434; Indels 12; Gaps 2;

QY 746 GGAAGTCTGCGTCCCATCTGTGATGTTGTGTGTCTATTCATGCGATGCTCATG 805  
Db 1414 GAATATCTGCGGAGATCAGATGCGCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1473  
QY 806 GTGTGGATGCTGCGCGCGAGATCAAGATTCGCGCAAGCGGCTCTCTCTCTCTCT 865  
Db 1474 GTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1533  
QY 866 CTCTCCATACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 925  
Db 1534 AGCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1593  
QY 926 ACTGAGGATCTCTGTTCTTCTGTCAGCGCTCAATGGGCAACTCTTGAAGCTCGGTGA 985  
Db 1594 GGAGCTGGGATCTGATCTTCAATCAACCAAGTGGGAACTCAGAAATGCCAGAGTGA 1653  
QY 986 TGTATCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1045  
Db 1654 TGGAAAGATGCTCCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1713  
QY 1046 ATGTTCTCTCTTACAGGCTTTCAGACATATATATATATATATATATATATATATATAT 1105  
Db 1714 ACTCTCTCTTCTTACAAATATTCACAACTGCTACAGGAGCACTCAATATGTCAC 1773  
QY 1106 ACTTTGACACCTTTTCAAGTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1165  
Db 1774 TGCACCAAGATGCTCCACAGCATCTTGTGCGGCTGTGTGTGTGTGTGTGTGTGTGT 1833  
QY 1166 CTGCGGTAGAACTCAAGCTCAGAGGTGGAGATGATGATGATGATGATGATGATGATGAT 1225  
Db 1834 ATGGCCAAATGAAAGCAAGATCAATTTGAAATGTG---GCAGACCAAGGGCCAGGCATT 1890

QY 423 AGGTGTCCTCGACAAATTCAGTCTCAAAAACCTGTTAAAGTTGGCA---ATTTCAC 479  
Db 284 AACTGGCGCTGGGGCCAGTACCAACCCCTCGGCTGCTCAGTCTCTGGAAGAGATCTGCC 343  
QY 480 CGGCATGAAGAAGTACTGGATACGCTCAAGCTCCCGGCTGGTGAATCCTCTTACT 539  
Db 344 CCGCGGTTAAAGTGTCCGCTATGCGCATCTGCATGATGACATCAATGAGCATGGCATGACT 403  
QY 540 ACGTGTGATCTGTGGTCTCTCTGTGTTATTAAGTATGAGTATCAGGCTCCAGCCCTTTC 599  
Db 404 ACAACACGATCATCGGATGGGGGGGTATTTACGTATGCGTCTCTCGCGCTATTAAC 463  
QY 600 CATGGCTATTGTCAAGCTAGTGGGAACTGCGTACCTCAGATCCAACTTGGCTG 659  
Db 464 CTGTGTGCGCATGAGACCAAGTCCGACAACGAGTGAACACCCGCTGTGCAACGGCGGTCA 523  
QY 660 CATCAATCAACAACATCAACATGGTATACAGAGTGCATCACTTACTTTTGGAAACAG 719  
Db 524 CCTCACTCAGACTATCTACTTCTTTACACGGGCAAGAGATCTTTGCAACGTATG 583  
QY 720 TTCTCCAAACAAAGCGATGSAATTAAGAAGGCTCTGGTCCCCATC---TGACTTGG 776  
Db 584 TATTGAGAGGACACAAGTCTAAAGCCGCTGGATGATGATGGGGCCGATCAAGCCGTCGGT 643  
QY 777 TGTGTGCTATTCAATCGCATGGCTCATAGTGTTCGAGTGTGTCGCCGAGAGATCAGA 836  
Db 644 CTCTGTGTGTGTTCCGGGGCTTCTTGTCTCTGCTACTATCTTCTTGTGAAAGATCAGGA 703  
QY 837 GTTCCGGCAAAAGCGGCTACTTCTCTCGGCTCTTCCATACGTTGTCAATGATCACTTAT 896  
Db 704 GTGCTGCAAAAGTGTGTGGGTGAGACAGCTTGGGCCCTTACGTGTGTGCTGATTTCTGC 763

Db 761 TGGCCGAGAGCGCTCAAGCTTCCAGGACGAGGAGGATACGCTACTACTTACCCAG 823

QY 957 AATGGCGCAACTCCTTAGCTGGGTGATGGTACTACGACGTACCGCAAGTGTCTTC 101

Db 824 AAGGACACAATTTGCAAAACTCTAAGGATATGGATTGACGGGCAATCCAGATTTCTTC 883

QY 1017 CTCTACCAATGTGCACCGGACCGATCATCATGTTCTCTCTTAAACAGGTTTCAGACATA 107

Db 884 CGCTGGTCCGGGTTGGGAACCTTACGCGGCTCTCCAGCTAACAAAGGTCACACACA 943

QY 1077 ATATCTACAGGAGATCGTTGGATTGTTAGCACTTGGACACCTTACAAAGTTCTGTGCT 113

Db 944 ACTGCTACAGGAGACGCGCTCATCATCTTTTCTATCACTGTTTGACACAGCTCTTCGTC 100

QY 1137 GGTGCAGATCTTGGGTATCCTTGGTAACTCCGCGTACGACTCACTCAAGTGGAG 119

Db 1004 GTTTGCTATTTTTCGCTTTGGGGTACATAGGCGACAGTTCAACAC---AAGAGCATCG 106

QY 1197 ATGTGGTCGGTCTGACGGTACCACATCTTTCCTTCATTTCATACCTGATGCCATTGCCA 125

Db 1061 AAGAGGTTGGCTCGAAGGCCCTCGACAGCTGGTGTTCATGTTACCCCGAGGCATCGCA 112

Db 1121 CCATGACGGGCTCCGCTGTTCTGGGCATCATCTTCTCTATCTTATACCTTGGGAC 118  
 QY 1317 TCGGCATCAGCTGGCTCTGCTATGACTTTCAACATTTGGCATGGACGGCTCCAC 137  
 Db 1181 TTGACAGTACTTTTGGAGGCTTGTGAGGAGTCCACCGGCTCTTTGGAGCAATATCCTC 124  
 QY 1377 GTGT-----ACCCAGCGTCTACATGTCAGACGATGACCTGTCCTGGGTTCTCGCTT 142  
 Db 1241 GAGTGTTAGGCAGACATCGCGAAGATATTGTGGCTGTACTCTCTCTCATCTAATTT 130  
 QY 1430 GGACTTGT--TTACTGCACACCGGCTGCACATATATTCTTGACCTTGTGATCACTACG 148  
 Db 1301 GGGCTCTGCCACACCATATCGGTGGTGTATACCTCTGTAGACTCACTCAATGTGTATG 136

Db 1361 GCCCTGATGGCGATCTATTCGTGATTTCTGTGAGCGCTGCGGCTGTGCTGGGTG 1420  
QY 1548 ATGATTTGAGANCTGTGCTAGACATTAGTTCAATGTTGGTAAACACTGCTGCTT 1607  
Db 1421 ATGGCGTCGACCGGTTCTCTGAAGATGTGAGACATGCTGGGACACCCCTGGATGT 1480  
QY 1608 ACTGGGCTCTGCTGGGCGGTATCATCTGCTCTAATATGACGACTGTGCTCTACG 1667  
Db 1481 TCTGGAGGACCTTGTGCTTACATCATGCTCCGATTTCTGCTGCTGCTGCTGCT 1540  
QY 1668 CTCTTCTGCTCTTACACACCTGTGTGCGAGACAC 1705  
Db 1541 CCGTCTGCGACACGAGAGATGCTGCGCGGGAATAC 1578

RESULT 4  
US-10-198-846-11010  
; Sequence 11010, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steilman, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198, 846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11010  
; LENGTH: 4305  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2, 4278, 4279, 4280, 4281, 4282, 4283, 4284, 4285, 4286,  
; LOCATION: 4287, 4288, 4289, 4290, 4291, 4292, 4293, 4294, 4295, 4296,  
; LOCATION: 4297, 4298, 4299, 4300, 4301, 4302, 4303, 4304, 4305  
; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-11010

Query Match 6.58; Score 187.2; DB 15; Length 4305;  
Best Local Similarity 47.78; Pred. No. 6.5e-30;  
Matches 679; Conservative 0; Mismatches 728; Indels 15; Gaps 4;

QY 233 GAGCTGAAACCCGACAGATGATGATGTCACACAAATTAATTTGATGCTCTGC 292  
Db 522 GACCTTCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
QY 293 ATGCTACATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 352  
Db 582 ATGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641  
QY 353 GAGGAGGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 412  
Db 642 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 701  
QY 413 TACTACTAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 469  
Db 702 TTTTACATGAGAGCTGCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 761  
QY 470 TCATTTTACCGGCAATGAAGGTACTGATACGCTCAAGCTCCGCTGCTGCTGCTGCTGCT 529  
Db 762 AAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821  
QY 530 CTGCTTACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 589

Db 822 GCTTCTACTACAAACACATCATGAGCTGGGCGCTATACCTCATCTCTCTTACAG 881  
QY 590 GCCACTTCCATGAGGCTATTTGTCAGCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 649  
Db 882 GACAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 941  
QY 650 ACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 709  
Db 942 TTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1001  
QY 710 TTGAGAACACTTCTCA--ACAAGGATGAAATGGAAGAGGCTCTGCTGCTGCTGCTGCT 766  
Db 1002 ACCGCGACGCTCTGAGATCCACCGGCTTAAGGAGGCTCCAGAGGAGGAGGAGGAGG 1061  
QY 767 TGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826  
Db 1062 TGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121  
QY 827 GAGTCAAGAGTTCGCGCAAGGCGCTACTTCTGCGCTTCTCCATACGTTGCTGCTGCT 886  
Db 1122 GCGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1181  
QY 887 ATCACTTATTCATACACCAATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946  
Db 1182 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1241  
QY 947 GTCAGGCTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1006  
Db 1242 TTGAACCCATTTGCGAGAACTCTGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1301  
QY 1007 GTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1066  
Db 1302 ATCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361  
QY 1067 TTGAGCATATATATATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1126  
Db 1362 TTCAACAAACAACTGCTCAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1421  
QY 1127 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186  
Db 1422 TTGCTTTCGAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1481  
QY 1187 GAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1246  
Db 1482 GATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541  
QY 1247 GCCATTCGCAAAACATTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306  
Db 1542 GCGATAGCCAAACATGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601  
QY 1307 GTGCTGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1366  
Db 1602 AGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661  
QY 1367 GCGTTCACAGTGT-----ACCAACGCTATCATGCTGAGGAGGAGGAGGAGGAGGAGGAGG 1417  
Db 1662 GAGTTCACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1721  
QY 1418 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477  
Db 1722 TGTCTTCTTGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1781  
QY 1478 GATCACTAGCGTGAACATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537  
Db 1782 GAG 1841  
QY 1538 TTTGCTGATTTATGATTTGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1597  
Db 1842 TCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901  
QY 1598 ACTGCTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1639  
Db 1902 CCGGGGTGCTTCTGAGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943



RESULT 5  
 US-09-815-923-9  
 : Sequence 9, Application US/09815923  
 : Publication No. US20020197644A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Gill, Sarjeet S.  
 : APPLICANT: Ross, Linda S.  
 : APPLICANT: The Regents of the University of California  
 : TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A1  
 : FILE REFERENCE: 023070-093800US  
 : CURRENT FILING DATE: 2001-03-23  
 : NUMBER OF SEQ ID NOS: 20  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 9  
 : LENGTH: 4956  
 : TYPE: DNA  
 : ORGANISM: Manduca sexta  
 : FEATURE:  
 : OTHER INFORMATION: neurotransmitter transporter encoded by inebriated  
 : OTHER INFORMATION: gene  
 : US-09-815-923-9

Query Match 6.2%; Score 176.4; DB 11; Length 4956;

Best Local Similarity 48.0%; Pred. No. 1.5e-27; Mismatches 683; Conservative 0; Mismatches 706; Indels 33; Gaps 5;

QY 258 TATGTCACAAACATGTAATCTTGTATGTCGATCGATACATCCGTCGTTGGGTA 317  
 DB 634 TCTGGCCACAAAGATGATGCTGCTGCGTGGATATCCGTCGCGCCGCA 693  
 QY 318 ACGGTGGCGTTCCTTACATCCGATACGAAATGAGAGAGTCTTCTGTCGAT 377  
 DB 694 ACGGTGGCGTTCCTTACATCCGATACGAAATGAGAGAGTCTTCTGTCGAT 753  
 QY 378 ACGATCATGCTCTTACTGTCGCAAGCCTGTGATCTTACTAGAGTGTCTCTGAG 437  
 DB 754 ACTTCATCATCTCTTACTGTCGCAAGCCTGTGATCTTACTAGAGTGTCTCTGAG 813  
 QY 438 AATTCAGTCAAGAACTCTGTAAAGTTGGTC---AATTCACGGGCAATGAAGTA 494  
 DB 814 AGTATACAGCACACAGGCGCTATTTGGGCTCTATCAAAATTTGTCACCTTTTAAAGCG 873  
 QY 495 CTGATATACGTCAGTCGCGCGCTGCGGTATACATCTCTTACTAGTGTGATCTG 554  
 DB 874 CTGCTTAGCAAGTGTGTGATCTCTTAAATGTCACGATATAGCTGTGATATG 933  
 QY 555 GTCTCTGTCTTACTTACTGATGAGCTTCCAGGCACTCTTCCATGGGCTATTTGTC 614  
 DB 934 CATGGGCAATATATATTTTTCACATCATTTCAAAACGAAAGTACCTTGGCAAGTGTCT 993  
 QY 615 AGCCTGAGTGGAGAACTGCGTACCTCCAGATCCAAACTTGTGATCACTCAACACA 674  
 DB 994 CCAATCATGAGAACACAGACAAATGTGGGTTCACAAATCAACACATCGAAACGAAAGC 1053  
 QY 675 TCACCAATGATGACAGATGCTCACTCTTCTTTGAGACAGTCTTCCAAACAAAGC 734  
 DB 1054 GATACAAACACACCAC---GAAACATTTTTCGAGAAAGATATGAAATATGATG 1107  
 QY 735 ATGAATTTGAAGAGATGTCGTCGCGCCCATCTGATCTTGTGTTGTTATTCATCG 794  
 DB 1108 CTGCTATTTAGT---ATCTGTGTGATGCGGTGGGATTTGGACGCTTGTGTTG 1164  
 QY 795 CATGGCTCAGTGTGCGAGATGTCGCGCCGAGAGATCAAGATTTCCGCAAGCGGCT 854  
 DB 1165 CTGGGTGTATGTTATTTTCCCTATGGAAGAAACATTAATCTTACGCAAAAGTGTG 1224  
 QY 855 ACTTCGTCGCGCTTCCCATAGCTGTGATGATCACTTTATTCATCAGCAATCATCC 914  
 DB 1225 ATATCAACAAACACTGCAATTTCTGTGATATAGTCTTCTTGGAGCGTCTTGGACAC 1284

QY 915 TGCCGGGCTACTGAGCGCATCCGTCTTCTGTCACGCGCAATGGCGGAACCTCTG 974  
 DB 1285 TCGATGAGAGATGAGAGCTTGAAGTCTTCTTAAACAGACTGGGAACCTTTAAAC 1344  
 QY 975 AGCTGCGTATGATGACTCAGCAGTCAAGAGTGTCTCTGATGAGTGCACCG 1034  
 DB 1345 AGTCAGGCGCTGGGTCAAGCTGTGTCAGATCTTAACTATAGGAAATGATTCG 1404  
 QY 1035 GACCGATCATGATGTTCTCTTCAACAGGTTTCAACATTAATATCTACAGGATGCTT 1094  
 DB 1405 GGTGATATATATGTTCTCTTCAACAGGTTTCAACATTAATATCTACAGGATGCTT 1464  
 QY 1095 GGATTTAGCATTTGAGACCTTCAACAGTCTTCAACAGTCTTCAACAGTCTTCAACAG 1154  
 DB 1465 TAGCCGATCTCTGATCAAGCTTATCTAGTCTATAGTGGCATATTCACCTTGTGTA 1524  
 QY 1155 TCTTTGATACCTGCGGTACAGATCAACAGAGTGGAGATGTCGTCGTCGCG 1214  
 DB 1525 CCATCGGAATATGCGCTTGCAGAAATATCTCCGCTGAAGAT---GTCAATAGCGGACA 1581  
 QY 1215 GTACCATGCTTGTCTTCAATTCATACCTGATGCCATTGCCAATACTTCAACCTCAGC 1274  
 DB 1582 GTCCAGGCGCTATTTATTTGATGATACCCGCAAGTATAGCAAAATGCCACATCTCAGC 1641  
 QY 1275 TATTCGCGTCTGTTCTCCGATGATGTCGCGGTGATGCGCTCATCCGTCGCTC 1334  
 DB 1642 TATGGGAGATGCTTCTTCTTCAATGTTCTATGCTTGAATTAATATGCAATTTGCTA 1701  
 QY 1335 TGTATGACCTTTCACACATTTGGGATGAGAGCGGCTCC-----AC 1376  
 DB 1702 TAGTGAAGTGTGCTGACTTCAATACAAAGAGATTTCCGACATGATACGAAAGAC 1761  
 QY 1377 GTTACCAACCGCTCTACATGTCAGGATGACCTGTTCTTGGGTTCTCTGCTGACTTG 1436  
 DB 1762 TTGATATCATGAATTAATTAATCTGCTGTTGTCGCTGATATTAATGCGGACGTC 1821  
 QY 1437 TTTACTGACACCGGCTGACAAATATTTCTTCACTGATGATACATGATGATGAT 1496  
 DB 1822 CACACATATTCATACGCGGTATATATGATTCATGATTAATGATTAATGCGGCTGCTC 1881  
 QY 1497 TCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1556  
 DB 1882 TCAGTATTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941  
 QY 1557 AGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1616  
 DB 1942 GAAGACTGCTGAGAAATATTAACAAATGACAGGTCGCAACCATCGCTATCTTTCAT 2001  
 QY 1617 TCTGCTGGGCGGTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1658  
 DB 2002 TCTGTTGGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2043

RESULT 6  
 US-09-815-923-13  
 : Sequence 13, Application US/09815923  
 : Publication No. US20020197644A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Gill, Sarjeet S.  
 : APPLICANT: Ross, Linda S.  
 : APPLICANT: The Regents of the University of California  
 : TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A1  
 : FILE REFERENCE: 023070-093800US  
 : CURRENT FILING DATE: 2001-03-23  
 : NUMBER OF SEQ ID NOS: 20  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 13  
 : LENGTH: 3513  
 : TYPE: DNA  
 : ORGANISM: Manduca sexta







TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 OTHER INFORMATION: n=a,t,g or c  
 US-09-954-456-1630

Query Match 5.68; Score 160.6; DB 11; Length 2283;  
 Best Local Similarity 46.3%; Pred. No. 2,4e-24;  
 Matches 722; Conservative 0; Mismatches 789; Indels 48; Gaps 4;

260 TGGTCCACAAACATGTAATTCCTGATCTGCTGATCGCTACATCCGCTGGTTGGGTAC 319  
 181 TGGACACGCGAGATGACCTTATCATCTGCTGGTGGCTGCGGCTGCGGCTCGGTAC 240  
 320 GTGTGGCGGTCCCTTCATCGGTACCAAGATGAGAGAGTGGCTTCTGTCGCTAC 379  
 241 GTGTGGCGGTCCCTTCATCGGTACCAAGATGAGAGAGTGGCTTCTGTCGCTAC 300  
 380 GTCATCGTCTTTTACTTGTGGCAAGCGCTGTACTACTTGAAGTGTCTCGACAA 439  
 301 GTCTGATTCCTCCGTGGTGGAGGATCCCATTTTCTTCCGAAATCTCATGGGCGAG 360  
 440 TTGAGTTCAGAAACCTGTGTAAGTTGGTCAATTCACCGGCGCATGAAGTACTGA 499  
 361 TTGATGAGGCGCGGACATCAATGTGTGAACTCTGTCCCTATTCAAGGCTGGGC 420  
 500 TACGCTCAAGCTCCGCGCTGCTGATCATCTGTCTTACTACTGATGATGCTGCTC 559  
 421 TATGCTCCATGGATGTCTTCTACTGCAACACTTACTACATCATGATGCTGGCTGG 480  
 560 TGTCTGATTTACTTACTGATGATGATGATGATGATGATGATGATGATGATGATG 619  
 481 GGGTCTTATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
 620 GAGTGGAGAACTGCGTACCTGATGATGATGATGATGATGATGATGATGATGATG 679  
 541 ACCGTGAGACCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 680 AATGATACAGCAGTGTCTAAC-----TCTACTTTTG 712  
 601 TTGGCCAACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
 713 AGAAGATTTCCAAACAAAGCATGATGATGATGATGATGATGATGATGATGATGATG 772  
 661 GAGAACAACTCTTGAAGGCTCTCACAGGCTGAGGCTTCCAGAGCCCTCAACTGGAG 720  
 773 TTGCTGTGTGATTTATTCATGCGATGATGATGATGATGATGATGATGATGATGATG 832  
 721 GTTACCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 833 AAGAGTTCCGGAAGGCGCTCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 892  
 781 AATATCAAGGGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
 893 TTATTCATCAACATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 952  
 841 CTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 953 CCTCAATGGGGAAGACTCTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1012  
 901 CCTGACTGTGCAAGCTGGGCTCCCTCAAGTGTGATGATGATGATGATGATGATGATGAT 960  
 1013 TTCTCTCTGACAGTGTGACCGGACGATCATGCTTCTCTCTCTCTCTCTCTCTCTCTCT 1072  
 961 TTCTCTTATGCTATGCGGCTGGGCGGCTGACAGCCCTGGGCGATCAATATGCTTCAAC 1020  
 1073 CATATATATCTACAGGATGCTGATTTTATGACATTTGACACCTTTACAACTTTCTTG 1132  
 1021 AATCACTGTCTACAGGATGCTGATTTTATGACATTTTATGACATTTTATGACATTTT 1080  
 1133 TCTGGGTGACAGTCTTGGTATCTTGTATCACTGCGGTACGATCAACTGACAGGTG 1192

1081 GCTGGCTTTGCTGCTTCTCCATCTGCGCTTCATGCGCACAGAGGCTGCTATATC 1140  
 1193 GAGATGTGGTGGCTGCTGGGCGGATGACAGCTTCTGCTTCAATTCATCCCTGATGCAAT 1252  
 1141 TCCAGGTGGGAGATAGGCGCTG---GCTTACCTTCACTGCTACCCACAGGCGCTC 1197  
 1253 GCCAAACATTCACACCTCAGCTATTCGCTGCTTCTTCTTCCATGATGATGCTGCTG 1312  
 1198 ACATGATGCTGCTGCGGCCACCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257  
 1313 GGTATGCGCTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1372  
 1258 GGTCTGACAGCCAGTTTGAAGTGTGAGAGGCTTCACTACCTGCGCTCTGATCTCTC 1317  
 1373 CCA-----CGTATCCACCGCTGTACATGTCAGAGGATGACCTGCTTTCG 1417  
 1318 CCGGCTCTCTACTACTTCCGCTTTTCAAGGAGATCTCCGCTGCTGCTGCTGCTGCTGCT 1377  
 1418 GGTTCCTGCTTGGACTGTTTACTGACACACCGGCTGGACAAATATTTCTTGAAGCTGTA 1477  
 1378 TGTCTGCTATGCTGCTCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1437  
 1478 GATCTACAGGTGACATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1534  
 1438 GACTACTACAGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497  
 1535 GTTCTGCTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1594  
 1498 GTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557  
 1595 AAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1654  
 1558 CGACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617  
 1655 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1714  
 1618 ATCTTCACTTCAACAGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677  
 1715 CCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1773  
 1678 CCGTGTGGGCTGAGGCGCATGGGCTGCGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1736

RESULT 9  
 US-09-919-039-377  
 Sequence 377, Application US/09919039  
 Publication No. US20030108871A1  
 GENERAL INFORMATION:  
 APPLICANT: Kaser, Matthew R.  
 TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
 FILE REFERENCE: PA-0035 US  
 CURRENT APPLICATION NUMBER: US/09/919,039  
 CURRENT FILING DATE: 2002-09-09  
 PRIOR APPLICATION NUMBER: 60/222,113  
 PRIOR FILING DATE: 2000-07-28  
 NUMBER OF SEQ ID NOS: 401  
 SOFTWARE: Perl Program  
 SEQ ID NO 377  
 LENGTH: 2919  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 OTHER INFORMATION: Incyte ID No. US20030108871A1 5834958CB1  
 US-09-919-039-377

Query Match 5.48; Score 154.8; DB 12; Length 2919;  
 Best Local Similarity 46.5%; Pred. No. 4.9e-23;  
 Matches 738; Conservative 0; Mismatches 812; Indels 36; Gaps 6;  
 221 GACACGACTTAAAGGCTGACCGCCAGAACGATATGATGCTCAACAACTTGAATTC 280  
 136 GACACGAGACGAGACCGAGGTGAAGATCGGGCCCAATGACCAACAGATGAGATT 195

OY	281	TTGATGTCCTGCATCGCTACATTCGCGTGCGTTGGGTAACGCTGCGCGCTTCCCTTCAAC	340
Db	196	GTGCTGTAGTGGCCGGGGAGATCATATTGGCTGGGGCAATGTCTGGAGGTCTCCATCTGC	255
OY	341	GCGTACCAAGATGAGAGAGTGTCTTCTCGTGCCATACGTACATGTTCTTTACTTGTGC	400
Db	256	TGCTACAAAAACGGAGGTGGAGCGCTCTTTCATCCCTACTTCATCTTCTTGTGTGCG	315
OY	401	GGCAAGCCTGTGTACTACTTAAGAGTGTGCTCTCGACAAATTCAGTTCAAGAAACHTGT	460
Db	316	GGCATCCCGGTGTTCTCTCGAAGGTGGCGTGTGGCCCAATACACGACCAAGGAGGTGTC	375
OY	461	AAAGTTT---GGTCAATTTCACCGGCGCATGAAAGGTACTGGATACGCTCAAGCTCCCGG	517
Db	376	ACAGCTGAGAGAAAGATGTGCCCTCTCCAGGGCATTTGGCTGGCACTGTGGTGTATC	435
OY	518	TGCGATTACATCTGTCTTACTAGCTGGTGATCTGTGGTCTCTGTCTATTACTTACT	577
Db	436	GAGTCATATTTGAATGTCTTACTACATCATCATCTTGGCTGGGCTCTTCTTACTGTTTC	495
OY	578	ATGAGCTTCCAGGGCACCTTCCATGGGCTATTGTGACGCTGAAGTGGAGAACTGCGTA	637
Db	496	AGCTCCTTCACCTCTGAGCTGCCCTGGACGACCTGCAACAACTTTTGGAAACAGAGCAT	555
OY	638	CCCTCAGAT-----CCAAACCTTCTCTCATAGTCAACAACATCAACATGATAC	688
Db	556	TGCAGGAGCTTTGGAACCACTCAGAGACCGGGACAGAGTACGCCCATTTGAGATTTTAC	615
OY	689	AGCAGTGTCAACTCTACTTTTGTAGAACAAGTTCCTCAACAAGCGATGATGAAGA	748
Db	616	TCACCTTGATGAGATTTCTGGAGAGAGAGATCTGGGCAATCACTCCGGGATCCA---T	672
OY	749	GGTCTCGGTGCCCCCATCTGTGTACTTGGTGTGTCTATTATCATGCAATGGCTCAATGG	808
Db	673	GACCTGGGCTCCCTGCGTGGGAGCTGGGCCCTGTCCCTCTGCTGTGCTCTGGGTACTGCG	732
OY	809	TTGGAAGTCTGCGCCCGAGAGTCAAGAATTCCGGCAAAAGCGGCTACTTCTCGCGCTC	868
Db	733	TATTCTGCACTCTGGAAAGGGGGTCAAGCTCCACAGGCAAGGTGGTTATTTCACACCCAG	792
OY	869	TTCCCATCGTGGTCATGATCACTATTATTCACCAACAATCATCCTCCCGGTCTACT	928
Db	793	TTTCCGTACTGATGCTGTGTCTATTTGTGTGATCAGAGTGTCACTCCCTCCGAGACCTAC	852
OY	929	GAGGCACTCCTGTTCTTCTGTGCACGCGCTAATAGGGGAAACTCTTGAAGCTCGGTATGAG	988
Db	853	CAGGGCATCATCTACTACTTGAAGCAATTTGTTCCGCGCTCAGAGACCTCAGAGTGTGG	912
OY	989	TACTCAGAGTCAAGCAAGTGTCTTCTCTGTGACATGTGTACCGGACCGATCATATG	1048
Db	913	ATGATGCGGGGACCCCAAGATCTTCTTCCCTTGGCATCTCGACAGGGGTGCTGCACACC	972
OY	1049	TTTCTCTCTTCACAGGTTTCAGACATATATATCTTCAGAGGTGCTGTGATTGTTACGAT	1108
Db	973	CTGGGCACTCTACAAACATATACACAACAACCTGTCAAGAGCTGATGCCCTCTGTCTC	1033
OY	1109	TTGACACCTTTTACAAAGTTCTTGTCTGSGGTGCACGATCTTCGGTATCCTTGTAACTGC	1166
Db	1033	CTGAACATGCTCACACAGCTTGTGTGGCTGGTGTGTCTCTCCATCTCTGG---CTTC	1088
OY	1169	GCGTACGAACTCAACTCTCAGAGGTGGGAAATGATGGTGGGTGCTGGGGGTACAGTCTTCT	1222
Db	1090	ATGTGCCAAGACAAAGGGGTGCCATTTTCTGAAGTGGCCGAGTCAAGTCTCGGGGTGCC	1143
OY	1229	TTTATTTTCAACCTGATGCCATTTGCCAAAACATTCACACTCAGCTATTTCTCGGTGCTG	1288
Db	1150	TTTATCGCTTCCCAAGGCTGTGACTATGTATGCCCTTATCCACACTGTGGTCTCGCTG	1209
OY	1289	TTTCTTCCGATGATGTGGTGTGGGTATGTGGCTATCCGTGTGCTCTGCTATCAGCTTTC	1346
Db	1210	TTTCTTATCATGCTCATATATTCCTTGGGGGTGACACCGAGTTTGTCTGTGTGAGAGTGGCTG	1265

Db 782 TCGTACAAAACGAGGTGGAGGCTTCTCATCCCTACCTTCATCTTCTTCTGCTGC 841  
 QY 401 GGCAAGCCTGTGACTAGAGTGTCTCGGCAATTCAGTTCAAAACTGCT 460  
 Db 842 GGATCCCGGTGTTCTTCGAGAGTGGCTGGGCAATACACAGGAGAGTGC 901  
 QY 461 AAGATT--GCTCAATTCACGGCCATGAAAGTACTGATAGCTACAGCTGCGGC 517  
 Db 902 ACAGCCTGGAGAAATCTCCCTCTTCAGAGGCAATGGCTGGCATCTGTGTATC 961  
 QY 518 TGGGTTTACATCTCTTCTTCTACAGTGTGTATCTGTGTCTGTCTGTATCT 577  
 Db 962 GACTCATATTGTAATGTTCTTCTATCATCTCTTCTGCTGGCTCTCTCTCTCT 1021  
 QY 578 ATGAGCTTCAGAGCCACTCTTCTCATGGCTTATTTGTCAGCTGAGTGGAGAACTG 637  
 Db 1022 AGCTCTTCTACCTTCTGAGCTGCTGAGACCTGCAACAATTTGGAACAGAGCAT 1081  
 QY 638 CCTCAGAT-----CCACACTTCTGATCATCATCAACAATCACAATGTATCC 688  
 Db 1082 TGCAGGAGCTTCTGAAACCACTCAGAGCGGCAAGTGAACCCCATTTGAAATTTTACC 1141  
 QY 689 AGCAGTCTCACTACTTTTGGAGAACGTTCTCCAAACAACGATGGAATGAAGA 748  
 Db 1142 TCACCTGTCAATGGAATCTGGAGAGACAGAGTCTGGGCAATCACTGGGCAATCA--T 1198  
 QY 749 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808  
 Db 1199 GACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1258  
 QY 809 TTTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 868  
 Db 1259 TATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1318  
 QY 869 TTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928  
 Db 1319 TTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1378  
 QY 929 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988  
 Db 1379 CAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1438  
 QY 989 TACTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048  
 Db 1439 ATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1498  
 QY 1049 TTTCTCTTACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108  
 Db 1499 CTGGGCACTTACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1158  
 QY 1109 TTGGACACCTTACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1168  
 Db 1559 CTGAACAGTGGCCACAGCTTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1515  
 QY 1169 GCTTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228  
 Db 1616 ATGTCTTCCAGAGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1675  
 QY 1229 TTTATTTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1288  
 Db 1676 TTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1735  
 QY 1289 TTTCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348  
 Db 1736 TTTCTTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1795  
 QY 1349 AACACATTTGGAGTGGAGCGCTTCCACGCTGTACCCAC-----GCTCTAC 1393  
 Db 1796 GTGACAGCTTCCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1855  
 QY 1394 ATGTGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1453

Db 1856 CTCATCTCCACATCCGCTCATGCTGCTACCTGATAGGCTTTTCTGCTGCTACCGAGGC 1915  
 QY 1454 GGAACATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1510  
 Db 1916 GGGATGTACTATCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1975  
 QY 1511 TGGCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1570  
 Db 1976 CTGCTGATGTTTGAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2035  
 QY 1571 GACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1630  
 Db 2036 AACATTTGAGGACATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2095  
 QY 1631 ATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1690  
 Db 2096 CTGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2155  
 QY 1691 GTGTTGGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1750  
 Db 2156 AAGTACAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2215  
 QY 1751 TTTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1776  
 Db 2216 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2241

RESULT 11  
 US-09-917-800A-1429  
 ? Sequence 1429, Application US/09917800A  
 ? Patent No. US20020119462A1  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Mendrick, Donna  
 ? APPLICANT: Porter, Mark  
 ? APPLICANT: Johnson, Kory  
 ? APPLICANT: Castle, Arthur  
 ? APPLICANT: Elashoff, Michael  
 ? APPLICANT: Gene Logic, Inc.  
 ? TITLE OF INVENTION: Molecular Toxicology Modeling  
 ? FILE REFERENCE: 44921-5038-US  
 ? CURRENT APPLICATION NUMBER: US/09/917, 800A  
 ? CURRENT FILING DATE: 2001-07-31  
 ? PRIOR APPLICATION NUMBER: US 60/222,040  
 ? PRIOR FILING DATE: 2000-07-31  
 ? PRIOR APPLICATION NUMBER: US 60/222,880  
 ? PRIOR FILING DATE: 2000-11-02  
 ? PRIOR APPLICATION NUMBER: US 60/290,029  
 ? PRIOR FILING DATE: 2001-05-11  
 ? PRIOR APPLICATION NUMBER: US 60/290,645  
 ? PRIOR FILING DATE: 2001-05-15  
 ? PRIOR APPLICATION NUMBER: US 60/292,336  
 ? PRIOR FILING DATE: 2001-05-22  
 ? PRIOR APPLICATION NUMBER: US 60/295,798  
 ? PRIOR FILING DATE: 2001-06-06  
 ? PRIOR APPLICATION NUMBER: US 60/297,457  
 ? PRIOR FILING DATE: 2001-06-13  
 ? PRIOR APPLICATION NUMBER: US 60/298,884  
 ? PRIOR FILING DATE: 2001-06-19  
 ? PRIOR APPLICATION NUMBER: US 60/303,459  
 ? PRIOR FILING DATE: 2001-07-09  
 ? NUMBER OF SEQ ID NOS: 1740  
 ? SOFTWARE: PatentIn Ver. 2.1  
 ? SEQ ID NO 1429  
 ? LENGTH: 2028  
 ? TYPE: DNA  
 ? ORGANISM: Rattus norvegicus  
 ? FEATURE:  
 ? OTHER INFORMATION: Genbank Accession No. US20020119462A1 M95762  
 ? US-09-917-800A-1429

Query Match 5.2%; Score 147.6; DB 11; Length 2028;  
 Best Local Similarity 46.7%; Pred.No.1,4e-21;  
 Matches 699; Conservative 0; Mismatches 764; Indels 33; Gaps 6;



Db 546 CGGTTACGGTATCTGCTATTTGACGTTACATGACATTTTCTATATATGATCATCCG 605  
QY 556 TC-----TCGTCTGATTTACTTACATGACATTTCCAGGCCATCTTCATG 603  
Db 606 TCAGCCGCTATTTTGTATTTGTTACATTTCAAAAATTTGGGATTCGAAAGTTCCG 665  
QY 604 GGCATATTTGACGCTGATGGAGAACTGCGTACCCAGATTCACAACTTGCTGCATC 663  
Db 666 GGCCTATCTGGCAATCCGCTGGAAATACACGAGATCTCAGATGACCTCAGTGCAT 725  
QY 664 ACTCAACAATCACCATATGATGACGAGCTCAGCTCACTTTTGAAGACATCT 723  
Db 726 ATCTGAAATGGAGACACCATTCACCTCCGTCAGAGATATTTATTTACAAATCTCT 785  
QY 724 CCAA---CAAGGATGATGAAAGGAGTCCGTTGCCCATCTGCTACTGTT 780  
Db 786 TGAAGTTCAAAAATCAACAGATTCGATATCTTTGAGGTGTAATAAATCTCAATGGCAGT 845  
QY 781 GTGTCTATTCATGATGCTCATGCTGATGCTGAGTCCGCGCCAGAGTCAAGAGTTC 840  
Db 846 GTGCTACTGCTGATTTATATATGTTTACTTGTCTTTGGAAGGCTCCACAGTCTC 905  
QY 841 CGGCAAGCGCCCTACTCTCGCGCTCTCCCATACGTTTCATGATCATCTTATCAT 900  
Db 906 TGGAAAATTTGTTGGGTACTGCAACAGCTCCATATTTATTTCTAAGTATTTCTCTAT 965  
QY 901 CACCAATCATCTGCGCGGTCTCTACGCGCATCCGTCTCTGTCGACCGCTCAATG 960  
Db 966 AGCTGACTCTTCTTCTCTGAGACAAAGATGCTCTTATTTATGATGACACCGGATTT 1025  
QY 961 GCGCAACTCTTGAAGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1020  
Db 1026 CAGAAACTCAAGATCTGAGTATGCTGCGGTGCTGCTACACAGATTTCTTCTCAT 1085  
QY 1021 GACAGTGTACAGGACCCATCATCTTCTCTCTTCAACAGGTTTACAGATTAAT 1080  
Db 1086 TGGACAGGATTTGGGTGCTGCTGCTGAGACATTCATGATTTTAAATTAATG 1145  
QY 1081 CTACAGGATGCTTGGATTTGATGATGATGATGATGATGATGATGATGATGATG 1140  
Db 1146 CTATGCTGACGCGCTATCTATCTCATCTATCTATCTATCTATCTATCTATCTAT 1205  
QY 1141 CACGATCTTGGTATCTTGTGATGATGATGATGATGATGATGATGATGATGATG 1200  
Db 1206 TGTGATATCTATACCTGCTGATATGCTCTTCCACCAATTAACCATTAATAGAT 1265  
QY 1201 GGTGCTGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
Db 1266 AGTTGAGAACAGAGAGCTCTCTATCTTCACTGCTACCCCAAGCCCTGCGACAT 1325  
QY 1261 ATTCAACCTGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
Db 1326 GATTAACATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1385  
QY 1321 CTATATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
Db 1386 CTTCATCTTTGCTGATGAAAGCATTTTACAGGGATCTGATGATGATGATGATG 1445  
QY 1381 ACCCAACCTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434  
Db 1446 GTGCAAAAATGCAAAATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1505  
QY 1435 TGTATCTGACACGCGGTGACAAATATATCTTGAAGTCTGATGATGATGATGATG 1494  
Db 1506 TCCCGCTATACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1565  
QY 1495 ATTCTTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1554  
Db 1566 TCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1625  
QY 1555 GAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1614  
Db 1626 TGAATGATTTCAAAAAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1685

QY 1615 TCTCTGCTGG 1624  
Db 1686 AGTCTGCTGG 1695

RESULT 13  
US-09-843-598-3  
; Sequence 3, Application US/09843598  
; Patent No. US2002010944A1  
; GENERAL INFORMATION:  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Ranganathan, Rajesh  
; TITLE OF INVENTION: CESPT GENES, PROTEINS, AND MODULATORY  
; FILE REFERENCE: 01997/525002  
; CURRENT APPLICATION NUMBER: US/09/843,598  
; PRIOR FILING DATE: 2001-04-26  
; PRIOR FILING DATE: 2000-04-26  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2016  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-09-843-598-3

Query Match 5.1%; Score 146.8; DB 10; Length 2016;  
Best Local Similarity 46.3%; Pred. No. 2.1e-21;  
Matches 644; Conservative 0; Mismatches 722; Indels 24; Gaps 4;

QY 259 ATGGTCCAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 318  
Db 306 ATGGGCACTAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365  
QY 319 CGTGTGCGGTTCCCTTTCATGCGGACAGAGATGAGAGAGTCTTCTGCTGCTGCTG 378  
Db 366 TATATGCGATTCCTCATGATGATGATGATGATGATGATGATGATGATGATGATG 425  
QY 379 CGTATCTGCTTCTTACTGCTGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438  
Db 426 TTTCAATTAATTAATGATGAGAGGACTTCCATGTTCTATATGAACTTGTACTGACA 485  
QY 439 ATTGATTAAGAAATCTGTTAAAGTTGTCAA---TTTCAACCGCATTAAGATGAT 495  
Db 486 ATTTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545  
QY 496 TGAATAGCTCAAGCTGCGCGCTGCGGTACATCTGCTTACTGATGATGATGATG 555  
Db 546 CGGTTACGGTATCTGCTGATTTTGGACGCTTCAATGACATTTCTATTAATGCGATAT 605  
QY 556 TC-----TCGTCTGATTTACTTACATGATGATGATGATGATGATGATGATGATG 603  
Db 606 TCAAGCCGCTATTTTGTCTATTTGCTACTTTCAAAAATTTGGGATTCGAAAGTCCG 665  
QY 604 GGCATATTTGACGCTGATGGAGAACTGCGTACCTCAGATTCACAACTTGCTGCATC 663  
Db 666 GGCCTATCTGGCAATCCGCTGGAAATACACGAGATGCTCAGATGACCTCAGTGCAT 725  
QY 664 ACTCAACAATCACCATATGATGACGAGCTCAGCTCACTTTTGAAGACATCT 723  
Db 726 ATCTGAAATGGAGACACCATTCACCTCCGTCAGAGATATTTATTTATTAACAAATCT 785  
QY 724 CCAA---CAAGGATGATGAAAGGAGTCCGTTGCCCATCTGCTACTGTT 780  
Db 786 TGAAGTTCAAAAATCAACAGATTCGATATCTTTGAGGTGTAATAAATCTCAATGGCAGT 845  
QY 781 GTGTCTATTCATGATGCTCATGCTGATGCTGAGTCCGCGCCAGAGTCAAGAGTTC 840  
Db 846 GTGCTACTGCTGATTTATATATGTTTACTTGTCTTTGGAAGGCTCCACAGTCTC 905  
QY 841 CGGCAAGCGCCCTACTCTCTGCGCTCTTCCCATACGTTGCTGATGATGATGATGATG 900



```

Db      906 TGGAAATTTGGGGTGAAGTCAACAGTCCATATATATTCATATTCCTTAT 965
Oy      901 CACACATATCTGCCCCGCTACGAGGSCATCTCTTCGACGCTCATG 960
Db      966 ACGTGGACTTTCTTCGAGCAAGATGCTCTATATATGACACGGATTT 1025
Oy      961 GCGCAACTCTGAGTCTGCTATGCTACGACATGACGAGTCTCTCTCT 1020
Db      1026 CGAATACTCAAGATCTGAGTATGCTGGCTGCTCTACAGATTTCTCTACT 1085
Oy      1021 GACAGTGCACCGGACCGATCAATGCTCTCTTACACGGTTCAACATATAT 1080
Db      1086 TGGACCAAGATTCGGGGTCTGCTGCTGAGAGTACATGATTTAAACAATACTG 1145
Oy      1081 CTACAGGAGATCTGATGATTTAGACCTTGGACACCTTACAGTTCTGCTGGGCTG 1140
Db      1146 CTATGCTGACCGCTCTACATATCTCATATTAAGTGCACGCTCATTTCTCGGATG 1205
Oy      1141 CACGATCTTCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db      1206 TGTGTATCTCTACCTTGGCTATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1265
Oy      1201 GGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db      1266 AGTTGGAGAACACGACGCTCTCTATATCTCTATGCTCTCTCTCTCTCTCTCTCT 1325
Oy      1261 ATTCACACCTGAGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db      1326 GGATTAAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1385
Oy      1321 CTATACCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
Db      1386 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Oy      1381 ACCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1434
Db      1446 GTCTCAAAATGCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1505
Oy      1435 TGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1494
Db      1506 TCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1565
Oy      1495 ATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1554
Db      1566 TCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1625
Oy      1555 GGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1614
Db      1626 TGATCAGTCTCAAAATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1685
Oy      1615 TCTCTCTCTG 1624
Db      1686 AGTCTGCTG 1695

```

# RESULT 14

```

: Sequence 2, Application US/09843598
: Patent No. US20020010944A1
: GENERAL INFORMATION:
: APPLICANT: Horvitz, H. Robert
: APPLICANT: Ranganathan, Rajesh
: TITLE OF INVENTION: CESTR GENES, PROTEINS, AND MODULATORY
: TITLE OF INVENTION: COMPOUNDS
: FILE REFERENCE: 01997/525002
: CURRENT APPLICATION NUMBER: US/09/843,598
: CURRENT FILING DATE: 2001-04-26
: PRIOR APPLICATION NUMBER: US 60/200,549
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2

```

```

: LENGTH: 2016
: TYPE: DNA
: ORGANISM: Caenorhabditis elegans
: US-09-843-598-2
Query Match      5.1%; Score 145.2; DB 10; Length 2016;
Best Local Similarity 46.3%; Pred. No. 4.7e-21;
Matches 643; Conservative 0; Mismatches 723; Indels 24; Gaps 4;
Oy      259 ATGGTCCAAACATTTGATTTCTGATGCTCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 318
Db      306 ATGGCAACTAAAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
Oy      319 CGTGGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 378
Db      366 TATATGGGATTTCCCATAGTATGCTACAAACAGCGTGGCTGCTTTCTTATTCCTA 425
Oy      379 GCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 438
Db      426 TTTCTATTTATGATGATGAGGACTTCCCATGTTCTATATGAACTTGTACTGAGACA 485
Oy      439 ATTCAGTTCAGAAACTCTGTTAAGTTTGCA---TTTACCGGCTATGAAGTATC 495
Db      486 ATTCATATGCTCAGGATGTTAGTATGAGAAAGGTGCTGCTGCTGCTGCTGCTGCTG 545
Oy      496 TGGATACGCTCAAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 555
Db      546 CGGTACGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 605
Oy      556 TC-----TCTCTCTATTTACTTATGCTATGAGCTTCCAGGCTCTCTCTCTCTCTCT 603
Db      606 TCAAGCCCTCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTG 665
Oy      604 GGTATTTGCTAGCTGATGCTGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
Db      666 GCGCTATGAGCAATCCGTGATACACCGAGATGCTGAGATGCTGAGATGCTGAGATGCTG 725
Oy      664 AGTCACAAACATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
Db      726 ATCTGAATATGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 785
Oy      724 CCAACAAACGATGATGATGAA--GAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db      786 TGAATTCAAATTCACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 845
Oy      781 GGTCTATTCATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 840
Db      846 GTGCTACTCTGCTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 905
Oy      841 GCGCAACGCGCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db      906 TGGAAATTTGTTGGGAGTCTGACGCAACGCTCCATATTTATTTATTTATTTATTTAT 965
Oy      901 CACCAATATCTCTGCGGCTGCTGATGAGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db      966 ACGTGGACTTTCTCTCTCTGAGCAAAATGCTCTATTTATTTATTTATTTATTTATTT 1025
Oy      961 GCGCAACTCTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db      1026 CGAATACTCAAGATCTGAGTATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1085
Oy      1021 GACAGTGCACCGGACCGATCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Db      1086 TGGACCAAGATTCGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1145
Oy      1081 CTACAGGAGATCTGATGATTTAGACCTTGGACACCTTACAGTTCTGCTGGGCTG 1140
Db      1146 CTATGCTGACCGCTCTACATATCTCATATTAAGTGCACGCTCATTTCTCGGATG 1205
Oy      1141 CACGATCTTCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db      1206 TGTGTATTTCTCTACACTTGGCTATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1265

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:24:46 ; Search time 72 Seconds  
(without alignments)  
1171.495 Million cell updates/sec

Title: US-09-991-458-2  
Perfect score: 3354  
Sequence: 1 MNDGVNGCFESSEPKMEPK.....TSRYKHLWYSITGAYRRNIN 633

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1577.5	47.0	641	22	AAE05101	Drosophila melanog
2	1523	45.4	593	22	ABBS9760	Drosophila melanog
3	1492	44.5	651	22	ABBS8403	Drosophila melanog
4	1472.5	43.9	639	22	ABW71178	Drosophila melanog
5	1472.5	43.9	639	22	AAE05104	Drosophila melanog
6	1414	42.2	629	22	ABBS4027	Drosophila melanog
7	1413	42.1	629	22	AAE00297	Drosophila melanog
8	1389.5	41.4	593	22	ABBS6992	Drosophila melanog
9	1389.5	41.4	593	22	AAE05103	Drosophila melanog
10	1086.5	32.4	797	19	AAW56376	Human Glyt-2 trans

11	1084.5	32.3	797	19	AAW56375	Human GLYT-2 trans
12	1084.5	32.3	797	19	AAW56369	Amino acid sequenc
13	1082.5	32.3	797	21	AAW82115	Human SC polypepti
14	1081.5	32.2	797	21	AAW56372	Human glycine tran
15	1080.5	32.2	797	19	AAW56372	Amino acid sequenc
16	1080.5	32.2	797	19	AAW56373	Human GLYT-2 trans
17	1080.5	32.2	797	19	AAW56373	Human GLYT-2 trans
18	1080.5	32.2	797	19	AAW56380	Human GLYT-2 trans
19	1080.5	32.2	797	19	AAW56381	Human GLYT-2 trans
20	1080.5	32.2	797	19	AAW56385	Human GLYT-2 trans
21	1080.5	32.2	797	19	AAW56388	Amino acid sequenc
22	1080.5	32.2	797	21	AAW56387	Human glycine tran
23	1080.5	32.2	797	21	AAW56387	Human glycine tran
24	1080.5	32.2	797	21	AAW56382	Human GLYT-2 trans
25	1080.5	32.2	797	21	AAW56382	Human GLYT-2 trans
26	1080.5	32.2	797	21	AAW56382	Human GLYT-2 trans
27	1080.5	32.2	797	21	AAW56382	Human GLYT-2 trans
28	1079.5	32.2	797	19	AAW56384	Human GLYT-2 trans
29	1079.5	32.2	797	19	AAW56384	Human GLYT-2 trans
30	1079.5	32.2	797	19	AAW56384	Human GLYT-2 trans
31	1079.5	32.2	797	19	AAW56378	Human GLYT-2 trans
32	1078.5	32.2	797	19	AAW56383	Human GLYT-2 trans
33	1078.5	32.1	797	19	AAW56383	Human glycine tran
34	1076.5	32.1	797	21	AAW56385	Human glycine tran
35	1076.5	32.1	797	21	AAW56385	Human glycine tran
36	1076.5	32.1	797	21	AAW56385	Human GLYT-2 trans
37	1073.5	32.0	797	19	AAW56386	Human GLYT-2 trans
38	1072.5	32.0	797	19	AAW56379	Human GLYT-2 trans
39	1071.5	31.9	797	21	AAW56320	Human glycine tran
40	1031	30.7	797	21	AAW56329	Human glycine tran
41	1031	30.7	797	21	AAW56329	Human glycine tran
42	1030.5	30.7	667	19	AAW57224	Rat prolactin transp
43	1027	30.6	637	18	AAW11704	High affinity Na+
44	1021.5	30.5	588	22	AAW11705	Human GABA transp
45	1019	30.4	642	21	AAW81944	Human GLYTLIKE pro

ALIGNMENTS

RESULT 1  
ID AAE05101 standard; Protein: 641 AA.  
AC AAE05101;  
DE 18-SEP-2001 (first entry)  
DT Drosophila melanogaster dmKSNF.  
DE Drosophila melanogaster.  
KW sodium/neurotransmitter family; SNF; biopesticicide; therapy;  
KW K+ coupled amino acid transporter.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200149848-A2.  
XX  
XX 12-JUL-2001.  
XX  
XX 28-DEC-2000; 2000WO-US35551.  
XX  
XX 30-DEC-1999; 99US-0173929.  
XX  
XX 15-MAR-2000; 2000US-0189399.  
XX  
XX 23-MAR-2000; 2000US-0191686.  
XX  
XX 23-MAR-2000; 2000US-0191687.  
XX  
XX 23-MAR-2000; 2000US-0191688.  
XX  
XX 23-MAR-2000; 2000US-0191695.  
XX  
XX (GENO-) GENOPTERA LLC.  
XX  
XX Kellerman KA, Keegan KP, Ebens AJ, Torrey J;  
XX

DR WPI: 2001-441879/47.  
 DR N-PSDB: AAD09681.  
 PT Novel invertebrate symporter cell surface receptor proteins and nucleic  
 PT acid encoding the protein useful as pesticide or drug target and to  
 PT identify compounds that have utility as therapeutics or pesticides.  
 PS Claim 18; Page 62-63; 71pp; English.

CC The invention relates to invertebrate symporter cell surface receptors  
 CC of the sodium/neurotransmitter family (SNF) and nucleic acid molecules  
 CC encoding such receptors. The SNF protein is useful for detecting a  
 CC candidate compound especially a putative pesticidal or pharmaceutical  
 CC agent that interacts with an invertebrate symporter cell surface  
 CC receptor protein or its fragment. Insect or worm genetically modified  
 CC to express protein of the invention is useful for studying invertebrate  
 CC symporter cell surface receptor protein activity by detecting the  
 CC phenotype caused by the expression or mis-expression of the protein in  
 CC the animal. Nucleic acids encoding the invertebrate receptor protein or  
 CC their fragments are useful as biopesticides. SNF nucleic acids are  
 CC useful for generating mutant phenotypes in an animal model or living  
 CC cells that are used to study the regulation of genes encoding the  
 CC proteins which are useful as pesticide or drug targets. The genetically  
 CC modified organisms or cells are useful in screening assays to identify of  
 CC pesticides or therapeutics and thus are useful in the identification of  
 CC new drug targets, therapeutic agents, diagnostics and prognostics  
 CC useful in treatment of disorders associated with ion channels. The  
 CC nucleic acid molecules are also useful as hybridisation probes.  
 CC The present sequence is *Drosophila melanogaster* (dm) K+ coupled  
 CC amino acid transporter of the SNF family which is referred as dmSNF.

XX Sequence 641 AA:

Query Match 47.0%; Score 1577.5; DB 22; Length 641;  
 Best Local Similarity 50.5%; Pred. No. 4e-156;

Matches 314; Conservative 90; Mismatches 183; Indels 35; Gaps 9;

QY 30 ANKALDNDITDPLEAEPERWWSNIEFLMSCIATSVGLGNVWFPEIAYONGGAP 89  
 DB 20 ATNAASSTK---TDAEPTAEFTNMGNGLEFLMSCSISVGLGNVWFPEIAYONGGAP 76  
 QY 90 LVRYVYVLLVGRKPYVLECVLGQFSRNSKYMS:SPARKGYSQAACGGITLSYVV 149  
 DB 77 LRIYIYVFLIGKPMYILEMIMGFTSQGIVKIMSVPGVGVGAFETICTIISTYSS 136  
 QY 150 ICGLCIYVLAMSFQATLPMAICQPEWENCVPSPDTLAASVNNITNGT----- 196  
 DB 137 LIALTLVYLFVSFQSELPMSYCRDEWNCVNSRP--QEVYDNLITGSLANESARNLSGI 194  
 QY 197 -----SSAQLYFLRTVLOOSDGIIEGGLCAPTIYLVLCFIAMLVFGVYANGVKS 246  
 DB 195 VANDETERKQSSSELYFLNVIYIEKIDISDGVDPMKLTALFVMMVYFLVIMGVKS 254  
 QY 247 SGRAAVFLALFPYVNMITLITITLLPGATDGIIFVTPOMAKLEIGVYSAVTOVFFS 306  
 DB 255 SGRAAVFLALFPYVNVLEVLIRAVTLLEGARDGILFLELPQMGMLNPTWKKEAVOCFFS 314  
 QY 307 LVYCTGPIIMESSYNGFRANIYDAMIYVTLDTFTSFLSGCTIFGLIGNAYLNSB-VG 365  
 DB 315 LNVGSGPIIMFASYNFHDGIYRDAMIYVTLDTLSTLGGITTFALIGNLAHMLQIENIR 374  
 QY 366 DVVAGAGTSLAFSYSDATAKTRQ--POLFSVLEFFLMASVLTGIGSSVALLSTNTLAMD 423  
 DB 375 DVV-RSGTGLAFISYDAISK--FOAVPOLFSVLEFFMLFVLGGSIVALLQSTIVTTICQ 432  
 QY 424 FPRVPTVYAMTSCGFLGLGYCTPGGQYILELDHYGFEFLFCALISELAFEMIT 483  
 DB 433 FPGMKYKVALTTSVCGFLMGVLYVTPGGQMLITLVDFGQIVYVITLAFELAGLVWY 492  
 QY 484 GLENLCLDIEMLGKTKGAYWRKLCWGVITPAIMTVFYALLASNNLVFGDNVYPTAG 543  
 DB 493 GLONFCDDIEFMGNRRVSLYWRVCMSEFTPMVMIIFISYMTIIEIKYSELY-FPEAN 551

QY 544 VSGYLMFLGTMFVPIGIGSLYKRTGSETIKKFKHSKPSWSPRRERKMEKA 603  
 DB 552 IAGWLFPAIGAQPPLMGWLTWISRHPOGTWKSLSKASLKDPSDWPANPEIRREWLEKN 611  
 QY 604 EAKALR-QKMNTSRVKHLWYSI 624  
 DB 612 QKAAGRATQKDTSLKGFWRKV 633

# RESULT 2

ABBS9760  
 ID ABBS9760 standard; Protein: 593 AA.

AC ABBS9760;

XX 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* polypeptide SEQ ID NO 6072.

KW *Drosophila*: developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS *Drosophila melanogaster*.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW.

XX WPI: 2001-656860/75.

DR N-PSDB: ABL03863.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -

XX Disclosure; SEQ ID NO 6072; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins  
 CC sequences (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

CC Sequence 593 AA;

XX Query Match 45.4%; Score 1523; DB 22; Length 593;  
 Best Local Similarity 51.0%; Pred. No. 1.9e-150;

Matches 301; Conservative 87; Mismatches 170; Indels 32; Gaps 8;

QY 62 MSCIATSVGLGNVWPPFIYONGGAFVYVYVLLVGRKPYVLECVLGQFSRNSYK 121  
 DB 1 MSCIATSVGLGNVWPPFIYONGGAFVYVYVLLVGRKPYVLECVLGQFSRNSYK 60  
 QY 122 VMSISPMKGTGYAAGCGYIISYVYVIGCLCLYVLAMSFQATLPMAICQPEWENCVPS 181  
 DB 61 IMSVVGFGVGVGAGFGLICISYSSLLALTLVYLFVSFQSELPMSYCRDEWNCVNS 120

QY 182 DPTLAASVNNITNGT-----SSAQLYFLRTVLOOSDGIIEGGL 218

```

Db 121 RP--OEYDNNLLTGTSLANESARNLSGIIVANDETEKLOSSSELYELNVIKEDIIDISGV 178
OY 219 GAPIYVYLCLFIAMLWGVAVGVARSSGKAAYFLALPPYVMTLFTTTIILPGADG 278
Db 179 GDPOMKLTALFVAVNVFLVIMGVKSSGKAAYFLALPPYVLTILIRATTLGADG 238
OY 279 ILFEVTPQAKLLELGVVYSAVTOVFESLTVCTGPIIMFSSYNGFRNHIYRDAMIYTLTD 338
Db 239 ILFELEPQWGLNPTVWKEAVVOCFSLAVSGPIIMFASYNRRDHGIYRDAMIYTLTD 298
OY 339 TPTSLSCCTTIGLIGNATAYELNSE-VGDVYAGAGTSIAFISYPAIAKTPO--POLFSV 395
Db 299 TLTSILGGITTFALLGNLAHNLQIENIRDVY-RSGTGIAFISYPAIAISK-FQAVPOLFSV 356
OY 396 LFFLLMSVILGIGSSVALSTFNTLAMDAPRPVPTVYMSAMTSCGFLGLVYITPGGOYI 455
Db 357 LFFFLMPLVIGISYVALOSTIYTTICDQFKGKRYKVALITTSVCGFLMGLVYITPGGOYI 416
OY 456 LELVDHYGTEFLVFCALSELAVGVIMYLENCLDIEFMLGKKTGAYWRLCQGYTPAI 515
Db 417 LTLVDYFGTYVVFILALFELAGIYVWYGLQNFCDIEFMCNRVSLYWRVQWSPFTPYM 476
OY 516 MTVVFYALLASNNLVFGDNVYPTAGYVSGTLMFLGNTFPIGTSILYKRTGTSE 575
Db 477 MIIETYSMTVETPIKYSLEY-FPEANIAAGWLFLAIGAOPPLMGWYISHHPOGYWK 535
OY 576 TIKKAFHSKPSWGPSPRREREMOFAKAKALR-OKMNTSRVKKLWYSI 624
Db 536 SIKASLKSDFRGNPANIPEIRREWVIFKNOKAQRATQDTSKLGFFWKV 585

```

### RESULT 3

ABBS8403 standard: Protein; 651 AA.

```

XX AC ABBS8403;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 2001.
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EM;
XX DR WPI; 2001-656660/75.
XX DR N-PSDB; ABL02506.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 2001; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of

```

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 44.5%; Score 1492; DB 22; Length 651;  
 Best Local Similarity 48.4%; Pred. No. 3,9e-147;  
 Matches 305; Conservative 96; Mismatches 191; Indels 38; Gaps 12;

```

OY 11 ESSEPKMEPRSSQSLPPANNKALDNDIDLEAEP--ERWVSNRIEIMACIATS 68
Db 38 ESSNKKTLSESGSAPRPHDS-----PSCVAGQPKRDRSMNDIERLMSCIALS 91
OY 69 VGLGVNWRPPTIAYONGGAFIVPYVYLLVGLKRPVYILECVLGFSSRNSYKWSISPA 128
Db 92 VGLGVNWRPPTIAYONGGAFIVPYVYLLVGLKRPVYILECVLGFSSRNSYKVDSPFI 151
OY 129 MKGTGYAAGGYYLISYVVICGLIYLAASPDATLPWATCOPEM-ENCYPSDPTLAA 187
Db 152 MKGIGYGOVLARIGVITYATLMALTRIFVDSFPTLPWSYCREMGTECLDSGPQERS 211
OY 188 SVNNITNGTSSAQLVFLRTVLOOSDIEGLGAPIMYLVLCFIAMLWGVAVGVKSS 247
Db 212 RATSILAGCGCHSPFHRNIIILREKASIDDGIGYPSKSLALAAVAVIAGIMFGVSS 271
OY 248 GKAAYFLALFPYVMTLFTTTIILPGATDGLFEVTPQAKLLELGVVYSAVTOVFESL 307
Db 272 GKASYFLALFPYVMTLFTTLPALTPGAFDGLVLEPQMKHLEPQVVAVTOVFESL 331
OY 308 TYCTGPIIMFSSYNGFRNHIYRDAMIVTLDFTSLSCTTIGLIGNATAYELN-SEVGD 366
Db 332 AICFQNIIMYASYNRGNHIYRDAMIVTLDFTSLSCTTIGLIGNATAYENNTDIALS 391
OY 367 VVGAGGTSIAFISYPAIAIK-TEFOPOLFSLFFLMSVYIGISSVALLSTFNTLAMDAPP 425
Db 392 VVN-GGPGIAFISYDAIAIKFKMLQLREVEFLMFLVIGISNNMGMSCHSTVAKDQNG 450
OY 426 RVPYVYMSAMTSCGFLGLVYCTPGGOYILEVDHYGTEFLVFCALSELAVGVIMYGL 485
Db 451 HLKNTVYVVGIAIVGFLGLVITPGGGLMLNVDYFVAVLAIPELVITAMIVGV 510
OY 486 ENLCIDIEFMLGKKTGAYWRLCQGYTPAIIMTVFFYALLASNNLVFGNYYPTAGYVS 545
Db 511 KRLCRDVERMIGIKTSLYRICWAVVTPMLMTLITLYLVLEPKYKD-YTYQSGVYVF 569
OY 546 GTLMFLGMTFVPIDIGFSLY-----KYRTGTFSEETIKKAFHSKPSMGPRSPRRR 596
Db 570 GWCL-----SAFGVGVLFMAIVAVRKQPSHLGLMAR-IRKAREPLPMKPSDPQILK 621
OY 597 EMMQFAE--AKALROKMTSRVKKLWYSI 624
Db 622 RYOLFVQEGNANALFRSS-----IMWKI 645

```

### RESULT 4

ABBS71178 standard: Protein; 639 AA.

```

XX AC ABBS71178;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40326.
XX XX
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.

```

XX MO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 XX 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI: 2001-656860/75.  
 XX N-PSDB; ABL15281.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 XX  
 XX Disclosure: SEQ ID NO 40326; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (AB101840-AB116175) and the encoded DNA  
 XX sequences (AB101840-AB116175) and the encoded proteins  
 XX (AB101840-AB116175).  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 XX Sequence 639 AA;

Query Match 43.9%; Score 1472.5; DB 22; Length 639;  
 Best Local Similarity 47.8%; Pred. No. 4.2e-145;  
 Matches 302; Conservative 97; Mismatches 200; Indels 33; Gaps 11;

QY 21 RSSQISIPANNKAL-DNIDDDLEAPPERAVWSNNITFLMSCTATSGLGNAWFFP 79  
 DB 19 RSTASTVEISTNSPALRNSDDDEAKVPERATWKGVEFLMSCTAMSGLNWVFFP 78  
 QY 80 IAYONGGAGFLVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 139  
 DB 79 TALDNGGAGFLVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 138  
 QY 140 CGYLLSYVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 192  
 DB 139 VFMLSTYVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 198  
 QY 193 TNGT-----SSAQLVFLRVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 243  
 DB 199 QNTTMSQNDRVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 258  
 QY 244 VKSSGAAFLVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 303  
 DB 259 VKSSGAAFLVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 318  
 QY 304 FRSYVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 362  
 DB 319 FRSYVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 378  
 QY 363 EGVGVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 420  
 DB 379 DIGSVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 436  
 QY 421 MDAPRVPYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 480  
 DB 437 RDRFPVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 496

QY 481 WYGLNCLDIEFLGKKTGAYWRLCGVITPAIMTVEFYALLASNNLEFGDNYPT 540  
 DB 497 WYGLNCLDIEFLGKKTGAYWRLCGVITPAIMTVEFYALLASNNLEFGDNYPT 555  
 QY 541 AGVYGLMFLGTFVPIGIFGLVYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 600  
 DB 556 WYGLNCLDIEFLGKKTGAYWRLCGVITPAIMTVEFYALLASNNLEFGDNYPT 613  
 QY 601 FKAFAKALROKMTSRKHLWYVYLLVGRPVYLLVGLGFSRNSRVKWSISPAKKGTYAAG 632  
 DB 614 ---HKEIENELTPKRGQIW---AAIKONI 637

RESULT 5  
 ID AAE05104 standard; Protein: 639 AA.  
 XX AAE05104;  
 XX 18-SEP-2001 (first entry)  
 XX  
 XX Drosophila melanogaster dmsNF3.  
 XX  
 XX Fruit fly: Invertebrate symporter cell surface receptor; dmsNF3;  
 XX sodium/neurotransmitter family; SNF; biopesticide; therapy.  
 XX Drosophila melanogaster.  
 XX WO200149848-A2.  
 XX 12-JUL-2001.  
 XX 28-DEC-2000; 2000WO-US3551.  
 XX 30-DEC-1999; 99US-0173929.  
 XX 15-MAR-2000; 2000US-0189399.  
 XX 23-MAR-2000; 2000US-0191686.  
 XX 23-MAR-2000; 2000US-0191687.  
 XX 23-MAR-2000; 2000US-0191688.  
 XX 23-MAR-2000; 2000US-0191695.  
 XX (GENO-) GENOPTERA LLC.  
 XX Kellerman KA, Keegan KP, Edens AJ, Torpey J;  
 XX WPI: 2001-441879/47.  
 XX N-PSDB; AAD09684.  
 XX  
 XX Novel invertebrate symporter cell surface receptor proteins and nucleic  
 XX acid encoding the protein useful as pesticide or drug target and to  
 XX identify compounds that have utility as therapeutics or pesticides -  
 XX  
 XX Claim 18; Page 67-68; 71pp; English.  
 XX  
 XX The invention relates to invertebrate symporter cell surface receptors  
 XX of the sodium/neurotransmitter family (SNF) and nucleic acid molecules  
 XX encoding such receptors. The SNF protein is useful for detecting a  
 XX candidate compound especially a putative pesticidal or pharmaceutical  
 XX agent that interacts with an invertebrate symporter cell surface  
 XX receptor protein or its fragment. Insect or worm genetically modified  
 XX to express protein of the invention is useful for studying invertebrate  
 XX symporter cell surface receptor protein activity, by detecting the  
 XX phenotype caused by the expression or mis-expression of the protein in  
 XX the animal. Nucleic acids encoding the invertebrate receptor protein or  
 XX their fragments are useful as biopesticides. SNF nucleic acids are  
 XX useful for generating mutant phenotypes in an animal model or living  
 XX cells that are used to study the regulation of genes encoding the  
 XX proteins which are useful as pesticide or drug targets. The genetically  
 XX modified organisms or cells are useful in screening assays to identify  
 XX pesticides or therapeutics and thus are useful in the identification of  
 XX new drug targets, therapeutic agents, diagnostics and prognostics  
 XX useful in treatment of disorders associated with ion channels. The  
 XX nucleic acid molecules are also useful as hybridisation probes.

CC The present sequence is *Drosophila melanogaster* (dm) SNF homologue  
 CC which is referred as dmsnf3.

SO Sequence 639 AA;

Query Match 43.9%; Score 1472.5; DB 22; Length 639;

Best Local Similarity 47.8%; Pred. No. 4,2e-145;

Matches 302; Conservative 97; Mismatches 200; Indels 33; Gaps 11;

```

OY 21 RSSQISLPPANNKAL-DNIDDTDLAEPPERWMSNNIEFLMGCIATSVGLGNWRPPE 79
DB 19 RSTASTVEISTNSPLRDNSDDQEAQVPEERATMGKVEFLMGLSCIASVGLGNWRPPE 78
OY 80 IAYNGGGAFLPYVIVLLVGVKPPVYLLGVQSSSNYSKVSISIPAMKLTGYAQAG 139
DB 79 TALNGGGAFLPYVIVLLVGVKPPVYLLGVQSSSNYSKVSISIPAMKLTGYAQAG 138
OY 140 CGYLLSYVYVIGCLLYLAMSFOATLPAICQPEWE-NCVSPD-----TLASVNNI 192
DB 139 VFMLSTYVYVYVIGCLLYLAMSFOATLPAICQPEWE-NCVSPD-----TLASVNNI 198
OY 193 TNGT-----SSAQLYPLRTVLAQSDGTEGGGAPIMYIVLCIFAMLMVGVARG 243
DB 199 QNTYMKSONDRVITSSSEWYFVKEVLEHEKPNIEEGIGLPEWELVGLFTAMSCVFFIRG 258
OY 244 VKSSGKAAYFLALPEPYVYVITLFTTIIIPGATDGIIEFYVPOWAKLLELGVWSAVYQV 303
DB 259 VKSSGKAAYFLALPEPYVYVITLFTTIIIPGATDGIIEFYVPOWAKLLELGVWSAVYQV 318
OY 304 FFSLTCTGPIIMSSSYNGFRHNIYRDAMIYTTIDTFTSISGCTIFGILCNLAELNS- 362
DB 319 FYSLSVCEGNIIIMSSFNKFGHNVHDAIYTGDTMTSLAGFTIFGILHAEICTD 378
OY 363 EVGDVVGAGTSLAFISYPAIAKTFQ--POLFSVLEFLAMSVLGISGVALLSTFNTLA 420
DB 379 DIGSVY-KGGAGLAFISTPDAIAK-EKMLPOLFSVLEFLAMSVLGISGVALLSTFNTAI 436
OY 421 MDAPRPVPTVYVMSATSCGFLGLGVYCTPGGOYILEVDHYGTFVLVFCALISLAVGF 480
DB 437 RDRPFNFQMGQCSLLIAVVSFFIGLMTTPGGQYMLLVDFGASMLLVGIALVLYIG 496
OY 481 WYIGLNICLIDIEFMLGKKTGAYWRLCMGVTPPALMTTFVFFYALLASNNLVFGDNYVPT 540
DB 497 WYIGDRLCKDIEFMLGKKTGAYWRLCMGVTPPALMTTFVFFYALLASNNLVFGDNYVPT 555
OY 541 AGYSGVYMLFLGTFVPIGIGSLYKRYRTGTFSEITIKKAFHRSKMGPRSPREREMQ 600
DB 556 WSTSIGMLITAFGLIQPLIMIVAVLRDPPGOTLGAKINGAFTPKKMGSPDPLREQY-- 613
OY 601 FKAEKALRQKMTSRVKNHLYSITGAYRNI 632
DB 614 -----HKEIENELTPKRGQIW-----AALKONI 637

RESULT 6
ABB64027
ID ABB64027 standard; Protein; 629 AA.
XX
XX ABB64027;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 18873.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX

```

PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABL08130.

PI New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions -

PS Disclosure: SEQ ID NO 18873; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 629 AA;

Query Match 42.2%; Score 1414; DB 22; Length 629;

Best Local Similarity 43.9%; Pred. No. 5,6e-139;

Matches 280; Conservative 113; Mismatches 209; Indels 36; Gaps 9;

```

OY 10 FESSEPKMEPRSSQISLPPANNKALDNIDDTDLAEPP-----RMYMS 55
DB 7 YOKLRNNNGAAGSREHGRGNSN-----DGISTVIYSABGEELTINCEASESSGQRDQMS 61
OY 56 NNIEFLMCIATSVGLGVNWRPPIAYONGGAPLVPIVYLLVGVKPPVYLLGVQSS 115
DB 62 RGVLEFLFSCIALSVGLGVNWRPPIAYONGGAPLVPIVYLLVGVKPPVYLLGVQSS 121
OY 116 SRNSKYVMSISIPAMKLTGYAQAGGCTLYVYVIGCLYLAWSFOATLPAICQPEW 175
DB 122 SRGCTRAFMDAPIMHNGIAYGOVYSTALATYIACIMALTIRLVASFEVLPWYICLVEM 181
OY 176 -ENCVPSPDLAASVNNITNGTSSAQLYELRTVLAQSDGI-EGGIGAPIMYIVLCIFAM 233
DB 182 GKSCVATGAT-AANDSSIVGVSSAELEFFQTGLREPSLDNDGIGTFSMDLVLCIATW 240
OY 234 LMFVGVVARGYKSSGKAYFLALPPYVYVITLFTTIIIPGATDGIIEFYVPOWAKLLEL 293
DB 241 VIIGTILSKGIRSSKASYFLALPEYVIMIVLLNRAVLPAGMAGIYVLELPOWSQLNLP 300
OY 294 GWYSAVYOVFFSLVCTGPIIMSSSYNGFRHNIYRDAMIYTTIDTFTSISGCTIFGIL 353
DB 301 HWYIAITQMFPSLAICGTLVYVMSFDFKNNKVDYIITTTIDSLTSIAGCTIFGIL 360
OY 354 GNLAYELNSEVDVVGAGTSLAFISYPAIAK-TFQDPLFSVLEFLMYSVLGISSVAL 412
DB 361 GNLAFETWKDISQVVGKAGLAFTSYDEALAFKRYLPQDLAVLEFFMLVYLGISNIGM 420
OY 413 LSTFNTLMDAFPRPYVYVMSATSCGFLGLGVYCTPGGOYILEVDHYGTFVLVFCAL 472
DB 421 ASAVVNVYKDRFTHLPWMLAVSASIIIGFLGLVYVMTGQGVFLVDFYCTFIALVLA 480
OY 473 ISELAVFMVIGLNICLIDIEFMLGKKTGAYWRLCMGVTPPALMTTFVFFYALLASNNLV 532
DB 481 IAEIYAVGMVITGYKICSDIEFMLNVKTSFWKRICMAIYVAGMLVLYVYMLFSEPLETY 540
OY 533 GDNVYVPTAGYVGYLMLGTFVPIGIGSLYKRYRTGTFSEITIKKAFHRSKMGPRSP 592

```

Db 541 -RGVQYPPVYVMGMIWGLVQLPFWALVTIYQCGKTEGSKFGLAMQPTAWMGPLOT 599  
QY 593 REREMWQKAEKALROK-----MNTSRVKKHLMYSTIG 626  
Db 600 Q-----KFEAYIIHRRKREADFKSPRGVLEDNITG 629

RESULT 7  
AAE00297  
ID AAE00297 standard; Protein: 629 AA.  
XX  
AC AAE00297;  
XX  
DT 13-JUN-2001 (first entry)  
XX  
DE Drosophila melanogaster bioamine transporter 1 (Bt1) homologue.  
XX  
KW Fruitfly; transporter protein; bioamine transporter 1; Bt1; screening;  
KW therapy; molecular transport disorder; biopesticide;  
KW invertebrate model.  
XX  
OS Drosophila melanogaster.  
XX  
FH Key  
FT Domain  
FT Location/Qualifiers  
FT 57..590  
FT /label= SNF domain  
FT /note= "Sodium neurotransmitter symporter family,  
FT specifically claimed in claim 1"  
FT 66..88  
FT /label= Transmembrane\_domain\_1  
FT 92..118  
FT /label= Transmembrane\_domain\_2  
FT 137..157  
FT /label= Transmembrane\_domain\_3  
FT 233..248  
FT /label= Transmembrane\_domain\_4  
FT 257..274  
FT /label= Transmembrane\_domain\_5  
FT 302..328  
FT /label= Transmembrane\_domain\_6  
FT 338..362  
FT /label= Transmembrane\_domain\_7  
FT 400..422  
FT /label= Transmembrane\_domain\_8  
FT 435..457  
FT /label= Transmembrane\_domain\_9  
FT 459..482  
FT /label= Transmembrane\_domain\_10  
FT 514..533  
FT /label= Transmembrane\_domain\_11  
FT 551..571  
FT /label= Transmembrane\_domain\_12  
XX  
PN WO200118178-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 08-SEP-2000; 2000WO-US24598.  
XX  
PR 09-SEP-1999; 99US-0153461.  
XX  
PA (GENO-) GENOPTERA LCC.  
XX  
PI Ebens AV, Keegan KP, Winslow JM;  
XX  
DR WPI: 2001-235196/24.  
DR N-PSDB: AAD03450.  
XX  
PT Drosophila melanogaster Bioamine Transporter 1 (Bt1) nucleic acid and  
PT protein, useful in screening assays to identify candidate compounds  
PT which are potential pesticide agents or therapeutics that interact with  
PT Bt1 proteins

XX  
PS Claim 2; Page 49-51; 53pp; English.  
XX  
CC The present sequence is a Drosophila melanogaster transporter protein  
CC homologue, bioamine transporter 1 (Bt1). The Bt1 is used in screening  
CC assays to identify candidate compounds which are potential pesticide  
CC agents or therapeutics that interact with Bt1 proteins. It can also be  
CC used to genetically modify metazoan invertebrate animals resulting in  
CC Bt1 expression or mis-expression. It can also be used in methods for  
CC identifying new drug targets, therapeutic agents, diagnostics and  
CC prognostics useful in the treatment of disorders associated with  
CC molecular transport across the membrane. The Bt1 nucleic acid or its  
CC fragments, such as an antisense sequence or double stranded RNA, may be  
CC used as a biopesticide to inhibit Bt1 function. The invertebrate model  
CC organisms such as Drosophila melanogaster are useful for rapidly carrying  
CC out large-scale systematic genetic screens and therefore for analysing  
CC expression and mis-expression of Bt1 protein.  
XX  
SQ Sequence 629 AA;  
XX  
Query Match 42.1%; Score 1413; DB 22; Length 629;  
Best Local Similarity 43.9%; Pred. No. 7.2e-139;  
Matches 280; Conservative 113; Mismatches 209; Indels 36; Gaps 9;  
QY 10 FESSEPKMEPRSSQISLPANNKKAALNDIDTDLAEPE-----RWWS 55  
Db 7 YOKLRMMQAGOSRREHPSN-----DGISTYIYSEBELTNCBAESESQORDWS 61  
QY 56 NNIEFLMNCIATSVGLGNVWRPFPIAYQNGGAFVLPYVTVLLVNGPVYVLECVLGQFS 115  
Db 62 RGVEFLFSCIALSVGLGNVWRPFPIALENGGAFLLPYVTVLLIGRPVYVLEVIIGQFS 121  
QY 116 SRNSVYVMSISPMKGGVYQAOAGCGIISYVYVIGCLCYLXLMSPQALPMAIQPEW 175  
Db 122 SRGCIKFDHAPIMKRGATIGQVYSTALATYICIMALTIRYLVASEVLPPTYCLVEW 181  
QY 176 -ENCVPDPTLAASVNNITNGTSSAQLFLRYLQOSDGI-EGSLGAPIVYLVLCFTAW 233  
Db 182 GKSCVATGAT-AANDSSIYGVSSAELEFQYTLREPSLDNDGLGTPSMDVLCLATW 240  
QY 234 LMFVGVYARGVKSSGAAYFLAFPPVYVMTLEITITIIILPGADGILFFVTPQAKLTEL 293  
Db 241 VIIGTILSKIRSSGRASYFLALFPYIMIVLLIRAVTLPAGAGVYVFLKPMOSDLNP 300  
QY 294 GWNYSATQVFSFLYTCGPIIMFSSYNGFRHNYIRDAWIVTLDFTFSLSGCTFGIL 353  
Db 301 HWYKALITQMFESLALCFGLVWYAFSFPDNKNVHKVDIITITIDSLTSLACIIFGIL 360  
QY 354 GNLAYLELNSEVGDVAGGTSIAFISYPAIAK-TPQPLFSVLFELMSVLGIGSSVAL 412  
Db 361 GNLAFTETKDISQVYKGGAGLAFISYPEALIAFKYLPQFLVLPFPMILVIGISNIGW 420  
QY 413 LSTFNTLADAFPRVFTVYVMSAMTSCSGFLIGVYCTPGQVITLVDHIGGFLVFECA 472  
Db 421 ASAVVNVVYDRFTHLPHMLAVASAITGFCGLVYVMPGQVFLNIVDEYGCFTIALVA 480  
QY 473 ISELAVENIYGLNCLDIEFMLGKKTGAYVWRCGVITPAITVFFYALASNNIYF 532  
Db 481 IAEILLAVGWITGVKRICSDIEFMLNKTSTYRICALIAYPGMLFVLVYMLFSEPLEY 540  
QY 533 GDNVYVPTAGVYSGVLMFLGMTFVPIGIGFSLYKRYGTFTSETIKAFHSKSPGRSP 592  
Db 541 -RGVQYPPVYVMGMIWGLVQLPFWALVTIYQCGKTEGSKFGLAMQPTAWMGPLOT 599  
QY 593 REREMWQKAEKALROK-----MNTSRVKKHLMYSTIG 626  
Db 600 Q-----KFEAYIIHRRKREADFKSPRGVLEDNITG 629

RESULT 8  
ABB06092  
ID ABB06092 standard; Protein: 593 AA.  
XX



Db	303	DTFTSLLSGVIIIFIGILGNLAHESGCTDIASVKA -GPGIAFTSYDADAAKRMPPQVPSL 361
QY	396	LEFLFMSVGLIGSGVALLSTFNTFLAMDAFPRVPVYMSAMTSCGFLIGLYCTPGGOYI 4555
Db	362	LEFLFMSVGLIGSGVAVGSCIMTVLKDQFVNVKLMIVVSLVLSIGFLVGLIYTPGGQHI 421
QY	456	LELVHDHGGTFELVFCFAISELAGVFMPIYGLFENLCLDIEPMLGKKYAWRLCMQVTPAI 515
Db	422	ITLMDPFGVTFVSLVNSAIFELLAVGMIYGTKKLCDAEAMTKSNYTRICWSIVPLV 481
QY	516	MTVVEFYALLASNNLVE-GDNY--VYPTAGY-VSGYLMFLGMTFVPPIGIFSLSYKRYTG 5711
Db	482	MLVILVYSLLTMRPLSYNGCEPFLYRYVGVGCVSGCI---IGOLF--YMAQYANFRQPKG 536
QY	572	TFTSETIKKAFHSKPSMGSPRSPPRRRMMQFKAALAKQK 611
Db	537	SLKSRINNSIKPHSDWGPSPDKKLMDOYF-----LNRK 570
RESULT 9		
ID	AAE05103	AAE05103 standard; Protein; 593 AA.
AC	AAE05103;	
DT	18-SEP-2001	(first entry)
DE	Drosophila melanogaster dmsNF2.	
KM	Fruit fly; Invertebrate symporter cell surface receptor; dmsNF2;	
KW	sodium/neurotransmitter family; SNF; biopesticide; therapy.	
XX	Drosophila melanogaster.	
XX	WO200149848-A2.	
PD	12-JUL-2001.	
PF	28-DEC-2000; 2000WO-US35551.	
PR	30-DEC-1999; 9GUS-0173929.	
PR	15-MAR-2000; 2000US-0189399.	
PR	23-MAR-2000; 2000US-0191886.	
PR	23-MAR-2000; 2000US-0191887.	
PR	23-MAR-2000; 2000US-0191688.	
PR	23-MAR-2000; 2000US-0191695.	
PA	(GENO-) GENOPTERA LLC.	
FI	Kellerman KA, Keegan KP, Ebens AJ, Torpey J;	
DR	WPI; 2001-441879/47.	
DR	N-PSDB; AAD09683.	
PT	Novel invertebrate symporter cell surface receptor proteins and nucleic acid encoding the protein useful as pesticide or drug target and to identify compounds that have utility as therapeutics or pesticides	
PS	Claim 18; page 65-66; 71pp; English.	
XX	The invention relates to invertebrate symporter cell surface receptors of the sodium/neurotransmitter family (SNF) and nucleic acid molecules encoding such receptors. The SNF protein is useful for detecting a candidate compound especially a putative pesticidal or pharmaceutical agent that interacts with an invertebrate symporter cell surface receptor protein or its fragment. Insect or worm genetically modified to express protein of the invention is useful for studying invertebrate symporter cell surface receptor protein activity, by detecting the phenotype caused by the expression or mis-expression of the protein in the animal. Nucleic acids encoding the invertebrate receptor protein or their fragments are useful as biopesticides. SNF nucleic acids are useful for generating mutant phenotypes in an animal model or living cells that are used to study the regulation of genes encoding the	



CC proteins which are useful as pesticide or drug targets. The genetically  
 CC modified organisms or cells are useful in screening assays to identify  
 CC pesticides or therapeutic agents and thus are useful in the identification of  
 CC new drug targets, therapeutic agents, diagnostics and prognostics  
 CC useful in treatment of disorders associated with ion channels. The  
 CC nucleic acid molecules are also useful as hybridisation probes.  
 CC The present sequence is *Drosophila melanogaster* (dm) SNF homologue  
 CC which is referred as dmsnf2.

XX Sequence 593 AA;

Query Match 41.4%; Score 1389.5; DB 22; Length 593;

Best Local Similarity 49.0%; Pred. No. 1.9e-136; Indels 19; Gaps 10;

Matches 284; Conservative 89; Mismatches 188; Indels 19; Gaps 10;

QY 39 IDDTLEAEPPERKVMNIEFLMCIATISVGLGNWRRFFIAYQNGGAFELPVYIVLL 98

DB 3 LESSELSPPDRKARNMNCSSLEFLMSCIALSVGLGNWRRFFIAYQNGGAFELPVYIVLL 62

QY 99 LVGKPVYVLECVLQGFSSRSYKWSISPAKKGIVQAAGCCYILSYVIGCLCYL 158

DB 63 VVGKPIYMEMLQGFSSRSYKWSISPAKKGIVQAAGCCYILSYVIGCLCYL 122

QY 159 AMSFOATLPMATCQPEW-ENCVPDPTLAASVNNITGSSAOLYFRTVLOQSDGIEG 217

DB 123 FDSFASLPMSPFCHEBMDGCVASGGQPLQGLSRFSSSTQYLRVILNETDSLEG 162

QY 218 LGADIVYLCLFETAMLVFVGVARGVKSSGKAAYFLAPYVYVMTLFTTITLPGATD 277

DB 183 IGYPGSLALMLGISMVLTLLIRGVKSSGKAAYFLAPYVYVMTLFTTITLPGATD 242

QY 278 GILFVPPOMAKLELVWYSAVYQVFFSLVTCGPIIMSSYNGFHHNYRDMAYVTL 337

DB 243 GVMFELPOMAKLELVWYSAVYQVFFSLVTCGPIIMSSYNGFHHNYRDMAYVTL 302

QY 338 DTFTSLSGCTFEGILGNLAELNS-EVGDVVGAGTSLAFISYPAIAR-TEQPOLFSV 395

DB 303 DTFTSLSGCTFEGILGNLAELNS-EVGDVVGAGTSLAFISYPAIAR-TEQPOLFSV 361

QY 396 LEFLMAYVLTGSSVALLSTFNTLAMDAPRVYVYVYVYVYVYVYVYVYVYVYVYV 455

DB 362 LEFLMAYVLTGSSVALLSTFNTLAMDAPRVYVYVYVYVYVYVYVYVYVYVYVYV 421

QY 456 LELVDHVGCTFLVFCALISELAGVWYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 515

DB 422 ITLMDHVGCTFLVFCALISELAGVWYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 481

QY 516 MTYVFFYVALLASNNLVF-GDNY--VYPTAG-VSGYLMPLGMYFVPIGIFSIXKRTG 571

DB 482 MLVILVYSLTMRPLSYNGOEFPLVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 536

QY 572 TFESEIKAFHSPKSPGPRSPREREMWQKAEKALROK 611

DB 537 SLKSRINNSIKPHSDMGSPDKLMDYOMF-----LRNK 570

RESULT 10

AAM56376

ID AAM56376 standard; Protein: 797 AA.

XX AAM56376;

XX 29-JUL-1998 (first entry)

XX Human GlyT-2 transporter protein R393G mutant.

XX Human, GlyT-2 transporter; glycine transporter; neurotransmitter;

XX activity; N-methyl-D-aspartate receptor; NMDA;

XX strychnine-sensitive glycine receptor; treatment; diagnosis;

XX nervous system disorder; myoclonus; muscle spasm; hyperactivity;

XX epilepsy; spasticity; head trauma; neuronal cell death;

XX multiple sclerosis; spinal cord injury; Huntington's disease;

XX amyotrophic lateral sclerosis.

XX Synthetic.  
 OS Homo sapiens.  
 XX Key location/Qualifiers  
 FH Misc-difference 393  
 FT /label="R393G  
 FT /note="Arg replaced with Gly"

XX WO9807854-A1.

XX 26-FEB-1998.

XX 20-AUG-1997; 97WO-US14637.

XX 20-AUG-1996; 96US-0700013.

XX (ALIX ) ALLELIX NEUROSCIENCE INC.

XX Albert VR, Borden LA, Kowalski LR, McKelvy JF;

XX WPI; 1998-169158/15.

XX Human glycine transporter - useful for treating nervous system

XX disorders, e.g. pain, myoclonus, etc.

XX Claim 1; Page -: 90pp; English.

XX The present sequence represents a human GlyT-2 transporter mutant  
 CC protein. GlyT-2 is a glycine transporter found predominantly in  
 CC the brain stem and the spinal cord. GlyT-2 regulates the levels of  
 CC glycine, a major neurotransmitter, that preferentially influences the  
 CC activity of N-methyl-D-aspartate (NMDA) receptors and  
 CC strychnine-sensitive glycine receptors. The GlyT-2 protein can be  
 CC used to treat or to diagnose a nervous system disorder or condition,  
 CC e.g. pain, myoclonus, muscle spasm or hyperactivity, epilepsy or  
 CC spasticity associated with stroke, head trauma, neuronal cell death,  
 CC multiple sclerosis, spinal cord injury, dystonia, Huntington's disease  
 CC or amyotrophic lateral sclerosis.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.

XX Sequence 797 AA;

Query Match 32.4%; Score 1086.5; DB 19; Length 797;

Best Local Similarity 38.2%; Pred. No. 1.9e-104; Indels 43; Gaps 11;

Matches 222; Conservative 107; Mismatches 209; Indels 43; Gaps 11;

QY 54 WSNNIEFLMCIATISVGLGNWRRFFIAYQNGGAFELPVYIVLLVGRPVYVLECVLQ 113

DB 194 WSKRDLFELSMVYAVGLGNWRRFFIAYQNGGAFELPVYIVLLVGRPVYVLECVLQ 253

QY 114 FSSRSNYSYKWSISPAKKGIVQAAGCCYILSYVIGCLCYLAMSFOATLPMATCQPEW 173

DB 254 FASOGPVYV 313

QY 174 EW-----ENCVPD-PTL-----AASVNNIT--NGTSSA-----QL 201

DB 314 PWMTECKDKTKLLDSCVLSDBPKIQIKNSTFCMAYPVYVYVYVYVYVYVYVYVYVYVYV 373

QY 202 YELRTVLOQSDGIF--GGGLAPYVYLCLFETAMLVFVGVARGVKSSGKAAYFLAPY 259

DB 374 YFKYFVLIKISAGIEYPEIG--WPLALCLFELSMVYAVGLGNWRRFFIAYQNGGAFELPVY 430

QY 260 VVMTLFTTITLPGADGILFVTPOMAKLELVWYSAVYQVFFSLVTCGPIIMSSYNGFHHNYRDMAYVTL 319

DB 431 VLVYVILIRGVTLPGAAGIYV 490

QY 320 YNGFHHNYRDMAYVTLTDTFTSLSGCTFEGILGNLAELNS-EVGDVVGAGTSLAFIS 379

DB 491 YNKRHHNCCYV 549

QY 380 YPDIAITPQOLFVYV 436

Db 550 YPEALTRLPSPFMAITFFMLTLGLDTPMFAITETIVTSISDEFPKILRTHKPVFTLGC 609  
 Qy 437 CSCGELLGLVYCTPGGQYILELVHYGTFVLFCALISELAGVEMVYGLNCLDIEFML 496  
 Db 610 CICEFIMGFPMITGGIMFQVDTYASVALVIAITFELGISTVIGLQPCEDIEMMI 669  
 Qy 497 GKKGAVRLCWGVITPAIMTTFEYFALLASNNLVFGDNVYYPAGVSGYIMLFLGMTF 556  
 Db 670 GFQDNIEFKVCMAVEPTILFTFLICEFYQWEMPTYG-SYRYPNMSVGLMMLACSVIWM 728  
 Qy 557 VPIGIGSLYKRYRTGTFSETIKKAFHSPGSPRPRERE 597  
 Db 729 IPMFVTKMH-LAPGRFIERLKLVCSPQPDWGPFLAQHGE 768

RESULT 11  
 AAM56375  
 ID AAM56375 standard; Protein: 797 AA.  
 AC AAM56375;  
 XX  
 DT 29-JUL-1998 (first entry)  
 XX

Human GlyT-2 transporter protein I279N mutant.  
 DE Human; GlyT-2 transporter; glycine transporter; neurotransmitter;  
 KW activity; N-methyl-D-aspartate receptor; NMDA;  
 KW strychnine-sensitive glycine receptor; treatment; diagnosis;  
 KW nervous system disorder; myoclonus; muscle spasm; hyperactivity;  
 KW epilepsy; spasticity; head trauma; neuronal cell death;  
 KW multiple sclerosis; spinal cord injury; Huntington's disease;  
 KW amyotrophic lateral sclerosis.  
 KW  
 XX Synthetic.  
 OS Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 279  
 FT /label= I279N  
 FT /note= "Ile replaced with Asn"  
 XX  
 PN MO9807854-A1.  
 XX  
 PD 26-FEB-1998.  
 XX  
 PF 20-AUG-1997; 97WO-US14637.  
 XX  
 PR 20-AUG-1996; 96US-0700013.  
 XX  
 PA (ALLX ) ALLELIX NEUROSCIENCE INC.  
 XX  
 PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;  
 XX  
 DR WPI: 1998-169158/15.  
 XX  
 PT Human glycine transporter - useful for treating nervous system  
 PS disorders, e.g. pain, myoclonus, etc.  
 XX  
 PS Claim 1; Page -: 90pp; English.  
 XX  
 CC The present sequence represents a human GlyT-2 transporter mutant  
 CC protein. GlyT-2 is a glycine transporter found predominantly in  
 CC the brain stem and the spinal cord. GlyT-2 regulates the levels of  
 CC glycine, a major neurotransmitter, that preferentially influences the  
 CC activity of N-methyl-D-aspartate (NMDA) receptors and  
 CC strychnine-sensitive glycine receptors. The GlyT-2 protein can be  
 CC used to treat or to diagnose a nervous system disorder or condition,  
 CC e.g. pain, myoclonus, muscle spasm or hyperactivity, epilepsy or  
 CC spasticity associated with stroke, head trauma, neuronal cell death,  
 CC multiple sclerosis, spinal cord injury, dystonia, Huntington's disease  
 CC or amyotrophic lateral sclerosis.  
 CC note: the present sequence does not appear in the specification; it was

CC created using information provided.  
 XX  
 SQ Sequence 797 AA;  
 Query Match 32.3%; Score 1084.5; DB 19; Length 797;  
 Best Local Similarity 38.2%; Pred. No. 3e-104;  
 Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10;

Qy 54 MSNNIEFLMCIANSVSGGNWRPFPIIYONGGAFVPIYVILLVCKPYYIECVGQ 113  
 Db 194 MSSKIDFLISMVGAIVAGNWRPFYILFQNGGAFILPILMMLALACLPIFFLEVSIGQ 253  
 Qy 114 FSSRNSYKWSISPAKMGTYAQAAGCCYIISYVYVIGLCLLYLAMSFQATLPALICOP 173  
 Db 254 FASQGPVSWKAIPALQCCGIAMLINSLVIAIYVVIICYLTFLEFAFVSVLPMGSCNN 313  
 Qy 174 EW-----ENCVPSD-PTL-----AASVNNIT--NGTSSA-----OL 201  
 Db 314 PWNPECKDKTKRLDSCVIDSHPKIQIKNSTFCMTAPVNTWMNFTSQANKTFVSGSEE 373  
 Qy 202 YELRTVLQSDGIBEGIGAPIMYIVLCLFIAMLVFGVAVRGVSSGKAAYFLALFPYV 261  
 Db 374 YFKTFVLKISAGIERP-GEIRMPALCLFLAMVIVYASLANGIKTSGAVVFTATPEYV 432  
 Qy 262 MITLEFITTILPGATDGLIFFTVQMAKLELVGYVSAVTVQVFSITVCTPIIMFSSYN 321  
 Db 433 LVILLIRGVTLPGAGAGIWFITTPKWEKLTDAVWKDAQTQIFPSLSAAMGGLITLSSYN 492  
 Qy 322 GFRHNITVDAMIVTTLDTFTSFLSGCTIFGLICNLAYELNSEVDGVAGGTSIAFISYP 381  
 Db 493 KEHNHCYDTLIVCTNSATSIIFAGVFISVIGFANERKRNIEVADO-SPGIAFVYVP 551  
 Qy 382 DAIAKTFQPLEFSVLFPLMSVVLGIGSSVALLTFNTLAMPAPRPVYVYSAWT---CS 438  
 Db 552 EALIRLPSPFWAITFFMLTLGLDTPFAITETIVTSISDEFPKILRTHKPVFTLGC 611  
 Qy 439 CGFLLGLVYCTPGGQYILELVHYGTFVLFCALISELAGVEMVYGLNCLDIEFMLGK 498  
 Db 612 CEFIMGFPMITGGIMFQVDTYASVALVIAITFELGISTVIGLQPCEDIEMMIGP 671  
 Qy 499 KTGAVWRLCMGVITPAIMTTFEYFALLASNNLVFGDNVYYPAGVSGYIMLFLGMTF 558  
 Db 672 QPNIFKVCMAFVPEPTILFTFLICEFYQWEMPTYG-SYRYPNMSVGLMMLACSVIWP 730  
 Qy 559 IGIGFSLYKRYRTGTFSETIKKAFHSPGSPRPRERE 597  
 Db 731 IPMFVTKMH-LAPGRFIERLKLVCSPQPDWGPFLAQHGE 768

RESULT 12  
 AAM56369  
 ID AAM56369 standard; Protein: 797 AA.  
 AC AAM56369;  
 XX  
 DT 29-JUL-1998 (first entry)  
 XX  
 DE Amino acid sequence of a human GlyT-2 transporter protein.  
 XX  
 KW Human; GlyT-2 transporter; glycine transporter; neurotransmitter;  
 KW activity; N-methyl-D-aspartate receptor; NMDA;  
 KW strychnine-sensitive glycine receptor; treatment; diagnosis;  
 KW nervous system disorder; myoclonus; muscle spasm; hyperactivity;  
 KW epilepsy; spasticity; head trauma; neuronal cell death;  
 KW multiple sclerosis; spinal cord injury; Huntington's disease;  
 KW amyotrophic lateral sclerosis.  
 KW  
 XX Homo sapiens.  
 OS  
 PN MO9807854-A1.  
 XX  
 PD 26-FEB-1998.  
 XX

PF 20-AUG-1997; 97MO-US14637.  
 XX  
 PR 20-AUG-1996; 96US-0700013.  
 XX  
 PA (ALIX ) ALLELIX NEUROSCIENCE INC.  
 XX  
 PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;  
 XX  
 DR WPI: 1998-169158/15.  
 DR N-PSDB: AAV22906.  
 XX  
 PS Human glycine transporter - useful for treating nervous system  
 PT disorders, e.g. pain, myoclonus, etc.  
 XX  
 PS Disclosure; Pages 48-50; 90pp; English.  
 XX  
 CC The present sequence represents a variant of the human GlyT-2  
 CC transporter. The cDNA encoding GlyT-2 was generated by reverse  
 CC transcription PCR (RT-PCR) from human spinal cord or cerebellum poly  
 CC mRNA. GlyT-2 is a glycine transporter found predominantly in the brain  
 CC stem and the spinal cord. GlyT-2 regulates the levels of glycine, a  
 CC major neurotransmitter, that preferentially influences the activity of  
 CC N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine  
 CC receptors. The GlyT-2 protein can be used to treat or to diagnose a  
 CC nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm  
 CC or hyperactivity, epilepsy or spasticity associated with stroke, head  
 CC trauma, neuronal cell death, multiple sclerosis, spinal cord injury,  
 CC dystonia, Huntington's disease or amyotrophic lateral sclerosis.  
 XX  
 SQ Sequence 797 AA;  
 Query Match 32.3%; Score 1084.5; DB 19; Length 797;  
 Best Local Similarity 38.2%; Pred. No. 3e-104;  
 Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10;

QY 54 MSNNIEFLMCIATSVGLGNVWRPPFAAYONGGAFIPVYIVLLVKGPPYVIECVLQ 113  
 DB 194 WSSKLDFTLSMVGAVLGNVWRPPYLAFOGGAFLIPYIMLALAGLPFFLEVSIGQ 253  
 QY 114 FSSRNSYKWSISIPAMGTGVAQAAGCYILSYVVICGLVLYLAMSFOATLPMAICOP 173  
 DB 254 FASGGRPSYVKAIKPAIDLOGGCIAMILINSVLAITYNVICTTFLFRASFVSILPMGSCNN 313  
 QY 174 EW-----ENCVPSPD-PTL-----AASVNNIT--NGTSSA-----QL 201  
 DB 314 PMWTPECKDKTKLLDSCVSDHPKIQIKNSTFCMTAYPNVTWNTFSQANKTFVSGSEE 373  
 QY 202 YFLRTVYQSDGIEGGIGAPITWVLCFLTAMLVFEGVAVARGVSSKAAVFLALPYYV 261  
 DB 374 YFKYEVLIKISAGIEYP-GEIRWPLALCLFLAMVIVVASLAKGITSKVYVFTATFPYV 432  
 QY 262 MITLEFITTILPGATDGLFFVPPQMAKLELGVYSAVQVPSFVTCGPIIMFSYN 321  
 DB 433 LVILLINGVTLPGAGAGIWFITPKWEKLDATYMKAAQIIFPSLSAANGILTTLSYN 492  
 QY 322 GFRHNIVRDAMIVTTLDFTSFLSGCTIFSLGNLAVELNSEVDVAGAGTSLATISYP 381  
 DB 493 KFNHNCYRDLIVCTFSATISIFAGFVIFSGVIGEMANERKVNINVAADQ-CPGIAFVYVP 551  
 QY 382 DATAKTQPOLESVLFFLMSVLCIGSSVALSLTFPNLAMDAPFRVTVMSANT---CS 438  
 DB 552 EALRLPLSPFWALITFLMLTLGLDTPMTATETIVTSIDEPKRYLRTKRPVFTLCCV 611  
 QY 439 CGEFLVYCTPGGOVYLEVDHGTFLVLCALISELAGVFIYGLNCLDIEFMIG 498  
 DB 612 CFPIMGPMITOGGIVMFOVDVYTAASAYALVITAIFFLVGISYVYGQRCEDELMIGF 671  
 QY 499 KTGAYWRLCNGVITPAITVTFVYVALLASNNLVFGDNVYVPTAGYVSGYLMFLGMPFV 558  
 DB 672 QPNFMKVCWAFVPTILTFILCFSSFQWPEPMPTVG-SYRXPNNMSVLGIMLACSVIWP 730  
 QY 559 IGIGFSIYKRTGTGTFSTIKKAFHSKPSNGPRSRERE 597

DB 731 IMFYIKMH-LAPGRFIERLKLVCSPQDMPGFLAQHRGE 768  
 RESULT 13  
 ID AAY82115  
 XX AAY82115 standard; Protein: 797 AA.  
 AC AAY82115;  
 XX  
 XX 05-JUN-2000 (first entry)  
 DE Human SC polypeptide glycine transporter SEQ ID NO:2.  
 XX  
 KW Human; glycine transporter; SC6 polypeptide; screening; diagnosis;  
 KW identification; neuropathic pain; spasticity; myoclonus; epilepsy;  
 KW stroke; head trauma; spinal cord injury; dystonia; multiple sclerosis;  
 KW arytrophic lateral sclerosis; Huntington's disease;  
 KW Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP982400-A1.  
 XX  
 PD 01-MAR-2000.  
 XX  
 PE 27-OCT-1998; 98EP-0203615.  
 XX  
 PR 28-AUG-1998; 98GB-0018890.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Brown AM, Chapman CG, Gloger IS, Evans JR, Cairns W, Heldon H;  
 XX  
 DR WPI: 2000-258613/23.  
 DR N-PSDB: AA295536.  
 XX  
 PT New isolated polypeptides and polynucleotides, useful for screening for  
 PT (anti)agonists and for diagnosing a disease e.g. spasticity, epilepsy,  
 PT stroke, multiple sclerosis, Huntington's disease or Parkinson's disease  
 XX  
 PS Claim 4; Page 17-20; 27pp; English.  
 XX  
 CC The present sequence represents the human SC polypeptide glycine  
 CC transporter. The polypeptide is useful in a method for screening to  
 CC identify compounds which stimulate or inhibit the function of the  
 CC polypeptide and in a process for diagnosing a disease or a  
 CC susceptibility to a disease in a subject related to expression or  
 CC activity of the polypeptide e.g. treatment of neuropathic pain, other  
 CC pain, spasticity, myoclonus, epilepsy, stroke, head trauma, spinal cord  
 CC injury, dystonia, multiple sclerosis, arytrophic lateral sclerosis,  
 CC Huntington's disease or Parkinson's disease.  
 XX  
 SQ Sequence 797 AA;  
 Query Match 32.3%; Score 1082.5; DB 21; Length 797;  
 Best Local Similarity 38.3%; Pred. No. 4.9e-104;  
 Matches 222; Conservative 105; Mismatches 213; Indels 39; Gaps 10;

QY 54 MSNNIEFLMCIATSVGLGNVWRPPFAAYONGGAFIPVYIVLLVKGPPYVIECVLQ 113  
 DB 194 WSSKLDFTLSMVGAVLGNVWRPPYLAFOGGAFLIPYIMLALAGLPFFLEVSIGQ 253  
 QY 114 FSSRNSYKWSISIPAMGTGVAQAAGCYILSYVVICGLVLYLAMSFOATLPMAICOP 173  
 DB 254 FASGGRPSYVKAIKPAIDLOGGCIAMILINSVLAITYNVICTTFLFRASFVSILPMGSCNN 313  
 QY 174 EW-----ENCVPSPD-PTL-----AASVNNIT--NGTSSA-----QL 201  
 DB 314 PMWTPECKDKTKLLDSCVSDHPKIQIKNSTFCMTAYPNVTWNTFSQANKTFVSGSEE 373  
 QY 202 YFLRTVYQSDGIEGGIGAPITWVLCFLTAMLVFEGVAVARGVSSKAAVFLALPYYV 261  
 DB 374 YFKYEVLIKISAGIEYP-GEIRWPLALCLFLAMVIVVASLAKGITSKVYVFTATFPYV 432

```

QY 26 MILEFTTIIILPGATDGLIEFTPOWAKLLEGVYSAVQVFFSLVYCGPIITMESSYN 321
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 LVLLIRGLVTPGAGAGIIVFTTPWEXLITNATVWKADATIEFSLSAWGGILLTLESSYN 492
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 GFRHNITRDAMIIVTTLEDTFTSPLESCITFGILIGNLAYENSEVDVYAGAGTSLAFTSY 361
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 493 KFLNNCKRDTLLVTCNSATSIFFAGFVLFYSIYGEANRKNIENTVADQ_GPGIAFVYVP 551
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 382 DAIAKTFOPOLFSFEFLMMSVYLGISSVALLSTFNTLAMDAPFRVPTVYMSANT---CS 438
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 EALTRLPLSPFAIIFFMLTLTLTGDTMEATIEITVTSIDSEPKYLRTHKPVYTLGCCI 611
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 439 CGEILGLVYCTGCGGYILEVDHYGTLVLFCAISLGYFWIYGLENCLDIEFMGLK 498
Db 612 CFEIMEPMITGGGIMFQVDTVAASALVIAIAFELVIGISYVYGQRECEDEIEMWIG 671
QY 499 KTGAYVRLONGVITPAIMTTFVFAVLAASNNLVREDNVVYTPAIVSGYILMLPLGMYFP 558
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 672 QPPIEFKVCMAVATPPIILFTILCFEYQMEPTVYG-STYRPNMSWMLGMLTCLACSYTWP 730
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 IGIGFSLVYKRTGTESETIKKAFHSKPSMGSPRPRERE 597
Db :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 731 IMEVIKMH-LAPGRFIERKLACASQDPWGPFLAOHRE 768

```

RESULT 14  
AAB23323

AC AAB23323;

DT	23-FEB-2001 (first entry)
XX	Human glycine transporter type 2 protein mutant #5
DE	
YY	

KW Human glycine transporter type 2; Glyt2; muscle spasticity; myoclonus  
KW epilepsy; stroke; head trauma; multiple sclerosis; spinal cord injury  
KW dystonia; Huntington's disease; amyotrophic lateral sclerosis;  
KW mutant; mutein.

OS Homo sapiens.  
OS Synthetic.

PH	key	Location/Qualifiers
FT	Misc-difference	188
FT		/note= "wild-type Asn substituted by Asp"
yy		

PN W0200029564-A2

PD 25-MAY-2000

PF 09-NOV-1999; 99WO-US26534.  
YY

PR 12-NOV-1998; 98US-0191468.  
XY

PA (GLIA-) GLIATECH INC.

PI Gallagher MJ, Burgess LH, Brunden KR;

DR WPI; 2000-387780/33.

PT New nucleic acid encoding a human glycine transporter type 2 (Glyt2)  
PT useful for expressing Glyt2 which can then be used for detecting  
PT whether an animal has autoimmune antibodies against a glycine  
PT transporter -

PS Claim 4; Page -; 175pp; English

CC The present sequence is a mutant version of the protein sequence of the  
CC human glycine transporter type 2 (GLYT2). Glycine is involved in  
CC neurotransmission in the spinal cord and cerebellum, where glycine  
CC receptor inhibition results in pain transmission. The GLYT2 gene and

protein could, therefore, ~~not~~ be used to identify inhibitory compounds which prevent G12R activity, causing a decrease in muscle hypercontractility. They could be used to treat diseases or conditions associated with muscle contraction, including muscle spasticity (such as that due to epilepsy, stroke, head trauma, multiple sclerosis, neuronal cell death, spinal cord injury and dystonia), myoclonus, epilepsy, Huntington's disease, anisotropic lateral sclerosis and pain.

Note: This sequence is not shown in the specification, but is derived from the hGlyT2 wild type sequence shown in SEQ ID NO: 124 (AAMB03899).

SQ	Sequence	797 AA;
----	----------	---------

Query Match	32.2%	Score 1081.5	-DB 21	Length 797
Best Local Similarity	38.1%	Pred. No. 6.2e-104		
Matches 223	Conservative 106	Mismatches 118	Indels 39	Gaps 10

QY 47 EPERWVWSNNIEFLMSCATSIYGVGNWRPEPIAYONGGGAFIYPIYVILJCKPRYY 100  
D 187 EDARGNWSKKDLFIITSMVGIAGLGNWRPEPIYALONGGGAFIIPYIMLALGLPIEF 244  
QY 107 LECVLGOFSSRNSVKWWSISPAWKGGYAOAGCYIISYYVIGLCLLYLAMSFOATL 166  
D 247 LEVSLGQFPASQGVSVWKAPALOGGIMLLIISVLAIYVWIIICYTLFIPLFASFSVL 306  
QY 167 PAIAIQEPW-----ENCYPSD-PLT-----AASVNNIT--NGTSSA-- 199  
D 307 PMGSCNNPNWTECKDKRTKLLDSCYISDHPKIQIKNSTFCMTATPANTWNTSOAKNT 366  
QY 200 -----OLYFLRVLQOOSDIEGGGLAPMYLVLCIFIMLWYGVGVAWGSSGAAYFL 25  
D 367 FVSGSEERYFYVAKISAGIETP-GETRPPLALCLFLAMVIVYASLAGITSGKYVET 422  
QY 255 ALFPYVMTLPTFTTILIPGANDGILFEVTPQMAKLEBEGVWASVAVOFSFLVCTGPI 314  
D 426 AFEPYVVLVILIRGVTEPGAGAIMYFTPRMKTLDATYWKDAQTQIFISLSAMGCL 488  
QY 315 IMSSSINGRRHNIYRDAIVYTLUDLFTSLSCCTIFGLGNALVELNSEVDGVAGAGTS 374  
D 486 ITLSSYNNKHNHCYRDLIVTCTNSATSIJFAQFVFEVSI GFAMERKNVAINVADQ- GPG 544  
QY 375 LAEISYDAIAKTFQDFLSVLEFIMSMVSLVIGSSVALLTFPNTLAMPAPRPVPTVWMSA 434  
D 545 IAAVYVPALTRILPSPFAWIIFFMLLTLLGLDTRFAETIETIYVISIDSEPRKYLRTHKPV 604  
QY 435 MT---CSGFFLLGLVYCPRPGQYIIELVNDHVGFTLVFCAISELAGVFWIYGENLCLD 491  
D 605 FTLGCCICEFIMGEPMTITGGIYMQVDVTAASALVIAITAFELVGISSYVGIQRCE 666  
QY 492 IEPMLGKTKGATWRLCOWGIYPAIMTVEFFVALLASNNLVPGDNVYPTAGVSGYLMFL 551  
D 665 IEMMIOFQFNIMFKWCMAFVPTIITLFIICFSFYQMEPMYTG-SYRYNMSMVLGIMLTA 722  
QY 552 LGMTFVPIGIGSIVKYRRTGSESEIHKAFHKSPEMSGRSRRERE 597  
D 724 CSYIMLPIMFVILKHM-LABGRITERKULVCSQPRMGPFLAONHKE 768

RESULT 15  
AAW56372  
ID AAW56372 standard; Protein; 797 AA

AC AAW56372;

DT 29-JUL-1998 (first entry)  
XX  
DE Amino acid sequence of a human GLYT-2 transporter protein

KM Human; *glyc-2* transporter; glycine transporter; neurotransmitter activity; N-methyl-D-aspartate receptor; NMDA; strychnine-sensitive glycine receptor; treatment; diagnosis; nervous system disorder; myoclonus; muscle spasm; hyperactivity; epilepsy; spasticity; head trauma; neuronal cell death; multiple sclerosis; spinal cord injury; Huntington's disease;

KW amyotrophic lateral sclerosis.  
XX Homo sapiens.  
OS  
XX WO9807854-A1.  
XX  
XX PD 26-FEB-1998.  
XX  
XX PF 20-AUG-1997; 97WO-US14637.  
XX  
XX PR 20-AUG-1996; 96US-0700013.  
XX  
XX (ALX ) ALLELIX NEUROSCIENCE INC.  
XX  
XX PI Albert VR, Borden LA, Kowalski LR, McKelvy JF;  
XX  
XX WPI; 1998-169158/15.  
XX  
XX DR N-PSDB; AAV22909.  
XX  
XX PT Human glycine transporter - useful for treating nervous system  
XX disorders, e.g. pain, myoclonus, etc.  
XX  
XX PS Claim 1; Pages 55-56; 90pp; English.  
XX  
XX CC The present sequence represents a human GlyT-2 transporter protein.  
XX CC The cDNA encoding GlyT-2 was generated by reverse transcription  
XX CC PCR (RT-PCR) from human spinal cord or cerebellum polyA mRNA. GlyT-2 is  
XX CC a glycine transporter found predominantly in the brain stem and the  
XX CC spinal cord. GlyT-2 regulates the levels of glycine, a major  
XX CC neurotransmitter, that preferentially influences the activity of  
XX CC N-methyl-D-aspartate (NMDA) receptors and strychnine-sensitive glycine  
XX CC receptors. The GlyT-2 protein can be used to treat or to diagnose a  
XX CC nervous system disorder or condition, e.g. pain, myoclonus, muscle spasm  
XX CC or hyperactivity, epilepsy or spasticity associated with stroke, head  
XX CC trauma, neuronal cell death, multiple sclerosis, spinal cord injury,  
XX CC dystonia, Huntington's disease or amyotrophic lateral sclerosis.  
XX  
XX SQ Sequence 797 AA;  
  
Query Match 32.2%; Score 1080.5; DB 19; Length 797;  
Best Local Similarity 38.2%; Pred. No. 7.9e-104;  
Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10;  
  
QY 54 WSNNIEFLMSCIATSVGLGNVWRPFLAYONGGAFLLPYVLLVGRPYVLECYLQ 113  
DB 194 WSKLDFILSNVGYAVGLGNVWRPFLAYONGGAFLLPYVLLVGRPYVLECYLQ 253  
QY 114 FSSRNSYKWSISPAKGTGTAAGGGLSYVYICGLIYLAISFOATLPMALCP 173  
DB 254 FASGPPVSMKALPALOGCGIAMLITISVLAITYNVITCYLFLFASVSVLPWGCNN 313  
QY 174 EM-----ENCVPD--PTL-----AASVNNIT--NGTSSA-----QL 201  
DB 314 PMNTPCKDKTKLLDSCVSDHKIOIKNSTFCMTAYPNVTMYNFTSQANKTFVSSSE 373  
QY 202 YFLFTVLQOSDGLIEGGLAPILWVLCFLTAMLVFSGVAVGKSSGAAVFLALPPYV 261  
DB 374 YFKFVLIKISAGIEXP-GEIRMPALCLFLAMVIVYASLAKGIKTSKVYFFATPPYV 432  
QY 262 MITLFTITILPGATDILFFVTPQAKLLELGWYSANVOFFSLFVCTGPIIMSSYN 321  
DB 433 LVILLINGVTLPGAGAIWYFTIPKWEKLDATYWKDAATOIFFSLAAMGGTLTSSYN 492  
QY 322 GFRHNIRADAMIVTTLTFTSFLSGCTIFGILGNLAVELNSEVGVAGAGTSLAFISYP 381  
DB 493 KFRHNCTRDILIVTCTSATISIRFAGVIFSYIGFMANERKVNLENVADQ--GPGIAFYVP 551  
QY 382 DATAKTPOLFSVLYFLIMSVLGIGSSVALLSTFNTLADAPRPVTVYMSAMT---CS 438  
DB 552 EALTRPLSPFMALIFFLMLTLTGDMFATIFETIVTSISDEPKYLRTKHVPFTLGGCI 611  
QY 439 CGFLGLVYCTPGGQYVLELVNDRGTFVLFCALISLAVGFWYTGLENLCIDTEPMK 498

DB 612 CFFIMGFPMITOGGIWFLVDVYAASTALVIAIFELVGISVYGLQRECEDEIMKIGF 671  
QY 499 KTGAVERLCKGVTPTAINTVFEFFALLASNNLVGDNVYVPTAGVSGYLMFLGTFVP 558  
DB 672 QPNIEMKVCAPVTPITLITLFCFSFYQWEPMTYG-SYRIPNMSMVLGMLMLACSVIWP 730  
QY 559 IGIGFSLYKRYGTGFSETIRKAFHSKPSMGPRSPRRERE 597  
DB 731 IMFVKMH-LAPGRFIERLKVCSQPDMDGPFLLQHRGE 768  
  
Search completed: July 21, 2003, 09:29:29  
Job time : 74 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:24:45 ; Search time 43 Seconds

(without alignments)  
1415.188 Million cell updates/sec

Title: US-09-991-458-2

Perfect score: 3354

Sequence: 1 MNDQVNGGESESEPKMKPK.....TSRVKHLWYSITGAVRRNIN 633

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1076.5	32.1	799	2 A48716	glycine transport
2	1027	30.6	637	2 JH0674	L-proline transport
3	1011	30.1	599	2 S11073	gamma-aminobutyric
4	1008	30.1	599	1 ACRGTG	gamma-aminobutyric
5	999	29.8	598	2 A47368	gamma-aminobutyric
6	991	29.5	630	2 A47398	serotonin transport
7	981	29.2	630	2 S30604	neurotransmitter t
8	979	29.2	611	2 S46260	creatine transport
9	978.5	29.2	627	2 F46027	gamma-aminobutyric
10	973	29.0	627	2 JH0695	gamma-aminobutyric
11	972	29.0	635	2 S23431	choline transport
12	971.5	29.0	607	2 S19585	serotonin transport
13	971.5	29.0	635	2 G02095	creatine transport
14	971	29.0	635	2 JG2386	creatine transport
15	970.5	28.9	635	2 A46061	Na(+)-dependent cr
16	968.5	28.9	620	2 S46487	taurine transport
17	968.5	28.9	620	2 G01426	taurine transport
18	967	28.8	692	2 I57956	glycine transport
19	966	28.8	655	2 A46270	Na and Cl dependen
20	965	28.8	602	2 A45078	gamma-aminobutyric
21	964	28.7	614	2 A43390	gamma-aminobutyric
22	963	28.7	627	2 B44409	gamma-aminobutyric
23	959	28.6	602	2 A44409	gamma-aminobutyric
24	956	28.5	638	2 I77912	glycine transport
25	954.5	28.5	619	2 S29839	taurine transport
26	954.5	28.5	622	2 S63539	GABA/beta-alanine
27	953.5	28.4	621	2 I57939	taurine transport
28	949.5	28.3	620	2 A48980	dopamine transport
29	946.5	28.2	620	2 I57937	dopamine transport

30	940.5	28.0	614	2 A41757	betaine transport
31	936	27.9	615	2 S43285	norepinephrine tran
32	935	27.9	633	2 I58140	glycine transport
33	935	27.9	633	2 S23151	glycine transport
34	935	27.9	638	2 JH0673	glycine transport
35	934.5	27.9	619	2 S20346	dopamine transport
36	933	27.8	619	2 I59558	dopamine transport
37	932.5	27.8	614	2 S68236	betaine/GABA trans
38	929.5	27.7	590	2 A47194	taurine and beta-a
39	929.5	27.7	597	2 S65673	GABA transport pro
40	929	27.7	602	2 I55651	norepinephrine tran
41	920.5	27.4	617	2 S14278	dopamine transport
42	902	26.9	693	2 A41617	dopamine transport
43	867.5	25.9	579	2 D88551	protein T23G5.5 [1
44	861	25.7	615	2 T43330	catecholamine tran
45	820	24.4	632	2 T24405	hypothetical prote

## ALIGNMENTS

### RESULT 1

A48716  
glycine transporter GXYT2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 02-Jun-1995  
C:Accession: A48716  
R:Lin. Q.R.; Lopez-Corcuera, B.; Mandlyan, S.; Nelson, H.; Nelson, N.  
J. Biol. Chem. 268, 22802-22808, 1993  
A:Title: Cloning and expression of a spinal cord- and brain-specific glycine transport  
A:Reference number: A48716; MUID:94043045; PMID:8226790  
A:Accession: A48716  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-799 <LID>  
A:Cross-References: GB:I21672  
C:Genetics:  
A:Gene: GXYT2  
C:Keywords: brain; transmembrane protein

Query Match	32.1%	Score 1076.5;	DB 2;	Length 799;
Best Local Similarity	37.5%;	Pred. No. 1.1e-74;		
Matches	224;	Conservative 107;	Mismatches 228;	Indels 39;
			Gaps	10;
QY	35	ALDNDIDDDLEAPPERNNNSNNIEFIMSCITSVGLGVWMPFPPIAYONGGATLPYV	94	
DB	177	ATTTPEDQGDENKARGNWSKLPDITLSMVAAGLGWMPFPYLARONGGATLPYL	236	
QY	95	IYLLVGRPVYIECEVCOFSSRNSVKVMSISPAKMGTCGYAAGCGYLSYVYICGIC	154	
DB	237	MMALALAGLPFLEFLVSLGCFASQGPVSWKATPALQCGIAMLIIIVLYNNICTY	296	
QY	155	LYLLAMSFQATLPMAICQPEW-----EKCVPSD-PTL-----AASVNI	192	
DB	297	LEVLFASFVSVLPMGSCNNPMTPECKDKTKLLSCVIGDHPKIQIKNSTPCMAYPNL	356	
QY	193	T--NGTSSA-----OLYELRTVLOQSDGIEGGAGAPITWLYCLPFIAMLVGVVAV	242	
DB	357	TIWNTTSQANKTFVGSSEYEFYVVKISAGIEYR-GETRWLPCLFLAMIVYASLAK	415	
QY	243	GVKSSGKAAYFLALPEYVVMITLFTTILIPGATDGLIEFVTPQNAKLELGVMASVQ	302	
DB	416	GKSGTKVYVFPATPEYVVLVLLRGVTLPGAGAGIWFELIPKMEKLLDATVMDAATQ	475	
QY	303	VAFSLVCTGPIIMSSYVGFHNRIYRDAMIYTTDTFSLSCCTIGICLNLAYELNS	362	
DB	476	IFFSLSAMWGLITLSSYVKNFHNOCYRDLLIYTCNSATSIAPGEVIVSIOFMANERV	535	
QY	363	EVGDVVGAGGTSLAFISYDPAITAFQQLFVFLFIMMSYVGISSVALSTFPTLMD	422	
DB	536	NIENVADO-GPGIATVYVPEALTRLPISFPMAIIFFLMLITGLDMRTITITVTSISD	594	
QY	423	APRPVPTVYMSAMT--CSCGFLIGLVCTPGGOYILLEVHDYGGTFVLFCALISLAGV	479	

Db 595 EEPKRLTRKRVFTLGGCCICFFIFGPMITGGGILYMDVDTYASALVITALELVI 654  
QY 480 FWIYGLNLCIDIEFMKKGKAGVRLCWMGVTIPALMTVFFVLLASNNLVPEGDNVYP 539  
Db 655 SYVGLDRFCDDIEEMIGFOPNIFWKCMAVPTLITLFCFSFYQMEPTYG-SYRYP 713  
QY 540 TAGVSGYMLMLFGLMTFVPIGIFSLYKRYTGFSEETIKAFHSPKSPGPRSPRRE 597  
Db 714 NMSVVLGMLMACSVINIPIMFVTKM-LAPGRIFELKLVCSQPDMGPFLLAQHRE 770

## RESULT 2

L-proline transport protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 28-Feb-1997  
C:Accession: JH0674  
R:Freneau Jr., R.T.; Caron, M.G.; Blakely, R.D.  
Neuron 8, 915-926, 1992  
A:Title: Molecular cloning and expression of a high affinity L-proline transporter expe  
A:Reference number: JH0674; MUID:92265305; PMID:1350201  
A:Accession: JH0674  
A:Molecule type: mRNA  
A:Residues: 1-637 <EXE>  
A:Experimental source: forebrain  
C:Superfamily: gamma-aminobutyric acid transporter  
C:Keywords: glycoprotein; leucine zipper; phosphoprotein; proline transport; transmembr  
F:46-65/Domain: transmembrane #status predicted <TM1>  
F:73-93/Domain: transmembrane #status predicted <TM2>  
F:76-97/Region: leucine zipper motif  
F:117-137/Domain: transmembrane #status predicted <TM3>  
F:217-236/Domain: transmembrane #status predicted <TM4>  
F:242-264/Domain: transmembrane #status predicted <TM5>  
F:291-311/Domain: transmembrane #status predicted <TM6>  
F:323-345/Domain: transmembrane #status predicted <TM7>  
F:374-397/Domain: transmembrane #status predicted <TM8>  
F:424-443/Domain: transmembrane #status predicted <TM9>  
F:457-479/Domain: transmembrane #status predicted <TM10>  
F:500-519/Domain: transmembrane #status predicted <TM11>  
F:537-559/Domain: transmembrane #status predicted <TM12>  
F:1/Binding site: phosphate (Thr) (covalent) (by CAMP-dependent kinase) #status predict  
F:41/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:132/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:239/600/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predict

Query Match 30.6%; Score 1027; DB 2; Length 637;  
Best Local Similarity 39.8%; Pred. No. 5.3e-71;  
Matches 227; Conservative 90; Mismatches 236; Indels 18; Gaps 11;

QY 41 DTDLAE-PPERWVSNIEFLMSCIATSVGLNVRFPFLAYONGGAEFLVPIVLL 99  
Db 26 DVIDDDVFAADRGWMTKLDLLELSCIGYGLGNVWFPYRAYNGGAEFLVPELMLAI 85  
QY 100 VGRPVYLLCEVLOEFSRNSVXWSTSPAKKGTGYAAGCGYLLSYVVICGLTYLA 159  
Db 86 CGPLLEFLLELSCQFSSGLPLAWKISPLERGAALLLVGLVAITYNMILAVLYELF 145  
QY 160 MSEOALPMAICOPEN-ENCV-----PSDPLAASVNNITNGSSAOLYELRYL--QOS 211  
Db 146 ASLTSLNPMWHGCGMMWTEKLEHGRKDGNCALPL-NLSSTYSPSEVSRVLAHQGS 204  
QY 212 DGBGLGADITWLVLCFLTAMLVGAVGVKSSGKAAYFLALFPYVMTLFTTTII 271  
Db 205 QGI-GRGGEIRWNLCTLLLAWVIFLCILKGVSSGKVVYFATFPYLLILMLLVGVT 263  
QY 272 LPGAQGLIFEVYPOMAKLELGWYSAYVQVFESLVCGPPIIMSSYNGFRINIRDA 331  
Db 264 LPGAQGLIFEVYPOMAKLELGWYSAYVQVFESLVCGPPIIMSSYNGFRINIRDA 331  
QY 332 WITVTLDTFSLGCTIFGLNLNLAELNSEVDVGVAGTSLAFISYDATAKTPOQ 391  
Db 324 FYITLGNATISLAGFAIFSVLGMSELGVV-DQVAKAGPLAFIYIQAAMTMLPLSP 382

QY 392 LFSVLEFLMSVLIGSSVALLSTENTLAMDAP---RVPIVWSAMTSCGFLGLVYC 448  
Db 383 FWSLEFFMLTLGLDQFALETITVAVDEFFYYLPKRAVNSGLICVAMTGLGLT 442  
QY 449 TPGQYITELVDHVGTFVDFCAISBELAGVWYTGLENCLDIEFMKKGKAGVRLCW 508  
Db 443 TDGGMVNLVLLDDYSASFGLWVVITTCIAVTRVYGIQRFCDIHMMLGFKPGLYFRACW 502  
QY 509 GVTPPALMTYVFFALLASNNLVFGDNVYFTAGYVSGYMLFLGMTFVPIGIFSLYK 568  
Db 503 LFLSPALLALLVYSIVKYQPESEYG-SYRFPAMAEELGLMLGLSLMIPAGMLVAVLR- 560  
QY 569 RTGFSEETIKAFHSPKSPGPRSPREREM 599  
Db 561 EEGSLMERLQOASRPALDMGP-SLEENRTGM 590

## RESULT 3

gamma-aminobutyric acid transport protein - human  
N:Alternate names: GABA transporter  
C:Species: Homo sapiens (man)  
C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999  
C:Accession: S11073  
R:Nelson, H.; Mandiyan, S.; Nelson, N.  
FEBS Lett. 269, 181-184, 1990  
A:Title: Cloning of the human brain GABA transporter.  
A:Reference number: S11073; MUID:90353567; PMID:2387399  
A:Accession: S11073  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-599 <EXE>  
A:Cross-references: EMBL:X54673; NID:931657; PIDN:CA83484.1; PID:931658  
C:Superfamily: gamma-aminobutyric acid transporter  
C:Keywords: transmembrane protein

Query Match 30.1%; Score 1011; DB 2; Length 599;  
Best Local Similarity 35.4%; Pred. No. 8.3e-70;  
Matches 213; Conservative 118; Mismatches 229; Indels 42; Gaps 13;

QY 3 DGQVNGFSESEPEMERKSSQIS-LPPANKKALNDITDLEAEPPERWVSNIEFL 61  
Db 10 DQGI-----STEVSEAVVANDPKPTLVYKQKKAADLPDRDTWGRDPL 54  
QY 62 MSCIATSVGLGNVRFPFLAYONGGAEFLVPIVLLVGRPVYLLCEVLOEFSRNSV 121  
Db 55 MSCVGYALIGLGNVRFPFLAYONGGAEFLVPIVLLVGRPVYLLCEVLOEFSRNSV 114  
QY 122 VMSISPAKGTGYAAGCGYLLSYVVICGLTYLAMSFQATLPMAICOPEN-ENCV 179  
Db 115 VMKLAPMFKGVLAAVLSPLNITYIIVISMAIYLYNSPTTLPKKQCDNPMTDRCF 174  
QY 180 PSDPLAASVNNITNGSSAOLYELRTVLDQSDIEBGLAPITVYLCLFIAMLVFV 239  
Db 175 SN-----YSWNTNMTSAVEEERWMMHMTDLDRP-QDIRPLAITTAIAWILVFC 228  
QY 240 VARGVSKSGAAAYFLALFPYVMTLFTITLIPGADGLIFETVPOMAKLELVGYSA 299  
Db 229 IMKVGNTGKAVYTSATPIIMLITLIFRGVTLPGAEGILFTITPFRKLSDBEVLLDA 288  
QY 300 VTQVFSLVCTGPIIMSSYNGFRINIYRDAMVITLDTFTSLGCTIFGLGNLAYE 359  
Db 289 ATQIFFSYGLGLSLALGYSNPHNNVYRDSITVICINCSMTSMFAFVFSIYGMVAV 348  
QY 360 LNSVGVNAGAGSLAFISYPAIATTPQOLFVFLPMLMSVLIGSSVALLSTPNTL 419  
Db 349 TKRSIAD-VAASGGLAFVPEAVTQLPISPLMAILFSLMLMLGLDQSCFCEVGRITA 407  
QY 420 AMDAFPRV-----PVIYSAMTSCGFLGLVYCPTPGQYITLVDHVGTFV-LVFCALS 474  
Db 408 LVDEYPLLNRRRELFIAAV-CITSYLIGLSNITGGIYVFKFLDYISASGMSLLPLVFF 466







Db 347 VTNRPAD-VAASGPGALFLAIPRAYTQPLSPMSLTFSSMLMLDIDQFCVBEFIT 405  
QY 419 LADAPRV-PTVMSAMTSCGFLGLVYCTPGQYIYLEVDHYGFTL-VLECAI 473  
Db 406 ALVDEFKRLGRREIF-AMVCIVSYGLISNTQGLYFKLFEDYSAGSMLELVE 464  
QY 474 SELAGVWYIGLEMLCDIEFMLGKTKGAYWRLCMGVITPAIMTVEFYALLASNNLVEG 533  
Db 465 FERYISMCGVNVNFENFNIEMWGHKPCLMWKLCHSFTPIYGVGLFSLAQMKPLMG 524  
QY 534 DNYVYPAGYVGYLMLFLGTFVPDIPGIFSLKYRT-CTPSETIK 578  
Db 525 -SYTFKMGQGVGMFALSSMLIP---GYMGYFELTSGSLKQRLR 567

RESULT 6  
A47398  
serotonin transporter - human  
N:Alternate names: 5-hydroxytryptamine transporter  
C:Species: Homo sapiens (man)  
C>Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999  
C:Accession: A47398; S37688  
R:Ramamoorthy, S.; Bauman, A.L.; Moore, K.R.; Han, H.; Yang-Feng, T.; Chang, A.S.; Ganap  
Proc. Natl. Acad. Sci. U.S.A. 90, 2542-2546, 1993  
A:Title: Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cl  
A:Reference number: A47398; MUID:93211998; PMID:7681602  
A:Accession: A47398  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-630 <RMA>  
A:Cross-references: GB:L05568; NID:9291819; PIDN:AAA3492.1; PID:9291820  
A:Experimental source: Placental trophoblastic cell line JAR  
A:Note: sequence extracted from NCBI backbone (NCBI:128282, NCBI:128283)  
J. Leisch, K.P.; Molozin, B.L.; Estler, H.C.; Murphy, D.L.; Riederer, P.  
R. Neural Transm. 91, 67-72, 1993  
A:Title: Isolation of a cDNA encoding the human brain serotonin transporter.  
A:Reference number: S37688  
A:Accession: S37688  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-630 <LESS>  
A:Cross-references: EMBL:X70697; NID:936432; PIDN:CAA50029.1; PID:936433  
A:Note: the authors translated the codon GGA for residue 25 as Val and GAA for residue 5  
C:Genetics:  
A:Gene: GDB:SLC6A4; HTT  
A:Cross-references: GDB:134713; OMIM:182138  
A:Map position: 17q11.1-17q12  
C:Superfamily: gamma-aminobutyric acid transporter  
C:Keywords: transmembrane protein

Query Match 29.5%; Score 991; DB 2; Length 630;  
Best Local Similarity 37.0%; Pred. No. 3e-68;  
Matches 226; Conservative 97; Mismatches 242; Indels 46; Gaps 15;

QY 1 MNDQVNGGSESEPEKPKRSSQISLPANNKAALDNDIDTDLAE--PPERWWSNNI 58  
Db 38 VESGQISNGY-SAVPSGAGDDTRHSIPAT-----TLVAELHOGRETWCKV 86  
QY 59 EFLMSCATGSGVGNWRPFITAYONGGAFVYVYVLLVNGRPVYLLCGVGGSSRN 118  
Db 87 DFLASTYIGAVDGNWRFPIYCYQNGGAFLLPYITMALEFGSIPLEMYELAGQYHRNG 146  
QY 119 SVKVM-SISPMKGTGYAQAAGCYIISYVVICGLCYLYLAMSFOATLPWALQOPME- 176  
Db 147 CISIMWKICIFIGICVAILCIIFAYIASYNTIMANALYLLISSFDOLPWTSCKNMNT 206  
QY 177 -KCVSPDPLIASVNNIT---NGTSSAQLYFLTYLQ--OSDIEGGLGAPVYLVLCF 230  
Db 207 GNC-----TNFSESDITWTLSHSTPAEEYTRHVIQHSKGLQ-DLGISWQLACIM 260  
QY 231 IAWLWEGVYARBGVSSKAAVFLALFPYVMTLTFTITITILGADIGLIEFTYPOAKL 290

Db 261 LITVITESTMKGVTKGKVVWVATPPIIISVLYRGATLPGAMRGVLYLPKNOKL 320  
QY 291 LELCWYSAYTVQFESLTCTGPIIMESYNGEHNITRDAMIYTLDTFTSLSGCTIF 350  
Db 321 LETGWIDMAAQIFESLGPFGVYLAFAASYNKNENNCYQDALVTSVNCMTSEFVSQVIF 380  
QY 351 GIIGNLYELNSENSEGDVVGAGSTAFISPDIAKTFQQLFVEFFIMMSYIGSSV 410  
Db 381 TVLGVMEMKREDYSEVAKDAGBSLFTTYAALANNPASTFAITFLITIGLDSTIF 440  
QY 411 ALSTFNTLADAPRV-----PTVMSAMTSCGFLGLVYCTPGQYIYLEVDHYG 464  
Db 441 AGLEGVITAVLDEPHWAKRERFVAVITCFEGLVLTLE---CGAVVKKLEBYAT 497  
QY 465 TFLVFCALISELAGVWYIGLEMLCDIEFMLGKTKGAYWRLCMGVITPAIMTVEFYAL 524  
Db 498 GPVALTVALLAENAVSMFYGTQFCRDVKEMLGSPWFMKICVWASPLFLITISFL 557  
QY 525 LASNNL-VFGDNYVYPAGYVGYLMLFLGTFVPDIPGIFSLKYRTGTPSETIKAFHS 583  
Db 558 MSPQLRLF-QVNPYMSIILGICITSSFCIPYIAVRLT-TTGTFRERLIK----- 610  
QY 584 KPSWGPSPRE 594  
Db 611 -SITPETPE 619

RESULT 7  
S30604  
neurotransmitter transport protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S30604  
R:Mayser, W.; Beyer, H.; Schloss, P.  
FEBS Lett. 295, 203-206, 1991  
A:Title: Isolation of cDNAs encoding a novel member of the neurotransmitter transport  
A:Reference number: S30604; MUID:92111740; PMID:1765155  
A:Accession: S30604  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-630 <MAV>  
A:Cross-references: EMBL:X63995; NID:956779; PIDN:CAA5401.1; PID:956780  
A:Superfamily: gamma-aminobutyric acid transporter  
C:Keywords: transmembrane protein

Query Match 29.2%; Score 981; DB 2; Length 630;  
Best Local Similarity 37.2%; Pred. No. 1.8e-67;  
Matches 226; Conservative 94; Mismatches 241; Indels 46; Gaps 15;

QY 5 QVNGGSESEPEKPKRSSQISLPANNKAALDNDIDTDLAE--PPERWWSNNIEFLM 62  
Db 42 QISNGY-SAVPSISAGDEASHIPAI-----TLVAELRGERETWCKMDFL 90  
QY 63 SCATISYGLGNWRPFITAYONGGAFVYVYVLLVNGRPVYLLCGVGGSSRN 122  
Db 91 SVIYAVADIGNIRFPYICYQNGGAFLLPYITMALEFGSIPLEMYELAGQYHRNGCISI 150  
QY 123 W-SISPMKGTGYAQAAGCYIISYVVICGLCYLYLAMSFOATLPWALQOPME--NCV 179  
Db 151 WRKICPFKIGIACIATYIASYNTIMANALYLLISSLTDRLPWTSCNMTGNC- 209  
QY 180 PSDTTLAASVNNIT---NGTSSAQLYFLTYLQ--OSDIEGGLGAPVYLVLCFIAWL 234  
Db 210 ---TNFAQDNITWTLHSRSPABEFLRVHVDIHQSKGLQ-DLGTSIQGLTLCIVLIF 264  
QY 235 MWFVYARVYKSSGKAAYFLALFPYVMTLTFTITILPGATGILFVTPQAKLLELG 294  
Db 265 VIFESIMKGVTKGKVVWVATPPIIISVLYRGATLPGAMRGVYLYLPKNOKLLETG 324  
QY 295 VWSAYTVQFESLTCTGPIIMESYNGEHNITRDAMIYTLDTFTSLSGCTIFGIG 354  
Db 325 VWYDAQAQIFESLGPFGVYLAFAASYNKNENNCYQDALVTSVNCMTSEFVSQVIF 384

QY 355 NLAELNSENVDVVGAGTSLAFISYPAIAKTFOPOLSVLEFLMAGVIGISSVALLS 414  
DB 385 YAAEMENEDVEYAKADGSLLEFITYAEAIAMNPSTFFAIIFFLLITLIGDSTAGLE 444  
QY 415 TINTLAMDAPFPRVPT-----YMSAMTSCSGFLGLVYCTPGGQYILEVDHYGGTFLV 468  
DB 445 GYITAVLDEPHIMAKRREMFVLIIVITCVLSLLTL---TSGAVVAVLLEBYATGPAY 501  
QY 469 LFCALISELAGVMIYGLNCLDIEFPLGKKTGAWRICMGVITPAIMTTFEYFALLASN 528  
DB 502 LVALLEAVNAVMPFIGITFCSDVKEMLFSFGWNRICWVAISPLLEFLITCSFSLMSP 561  
QY 529 NT-VFGDNVYPTAGVSGYLMFLGMTPFVPIGIFSLKRYRGTSETIKRAHSPSPM 587  
DB 562 QLRLF--QYNYPHMSIVLGYCIGMSSVICIPIYIIYRLIS-TPGTAKENIR-----SI 612  
QY 588 GPRSPRE 594  
DB 613 TPEPTPE 619

## RESULT 8

S46260  
creative transporter - marbled electric ray  
C:Species: Torpedo marmorata (marbled electric ray)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Aug-1999  
C:Accession: S46260  
R:Gulmbal, C.; Kilmann, M.W.  
J. Mol. Biol. 241, 317-324, 1994  
A:Title: A creative transporter cDNA from Torpedo illustrates structure/function relationship  
A:Reference number: S46260; MUID:94334996; PMID:8057375  
A:Accession: S46260  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-611 <GUI>  
A:Cross-references: GB:X77306; NID:9541753; PIDN:CA54512.1; PID:9541754  
C:Superfamily: gamma-aminobutyric acid transporter

## Query Match

29.2%; Score 979; DB 2; Length 611;  
Best Local Similarity 35.1%; Pred. No. 2.4e-67;  
Matches 208; Conservative 103; Mismatches 235; Indels 46; Gaps 10;

QY 47 EPEEMWNSNIEFLMCLATISVGLGNWREFFIAYONGGAFIVPYVILLVGRPYV 106  
DB 32 EPGSVHTWSRQMDFTMSCVGFAYVGLGNWREFFIAYONGGAFIVPYVILLVGRPYV 91  
QY 107 LECVIGPFSRSNSVWMSISPAKMGTYAQAAGCYIIISYYVIGCLLYLAMSPQATL 166  
DB 92 LEISIGPMAKAGIINAMNIAFLFKGLGYASVYVFCNTYTLVLTWSSEFYLVOQSSPL 151  
QY 167 FWAICQPEMENCVPSPDPTLASVNNITNGTSSAQ---LYFLKTVLQOSDGIIEGIGAPI 222  
DB 152 FWASCNNTWNTAAACEAGANAS-TEIYPTAPAGSIYVOFMRERVLRLSSGL-GVYGEIG 209  
QY 223 WTVLYCLFAMLMVGVAARGVSKGAAYFLALFPYVMTLFTTTIILPGADGILFF 282  
DB 210 WELTLCIAITMMLVFCIMKGVKTSKAVYVATFPYIIVILLVLRGVTLEGAOAGIYY 269  
QY 283 VTPQAKLLELGVMSAVTQVFFSLVCTGPIIMSSYNGFRHNIRDAWITTLDTPTS 342  
DB 270 LQPMGKLGEOVWMDAQTFEFSYALGTLTALGSLNOLHNDCKRAFIILSVNSATS 329  
QY 343 FLSCGTITGILGNLAELNSENVDVVGAGTSLAFISYPAIAKTFOPOLSVLEFLMMS 402  
DB 330 FFAGLVASISILGFMAVEGVDI-SVAESGPGIAFIAPKAVITLMPFQVAVLEFIMLL 388  
QY 403 VLGIGSSVALLSTFNTLMDAPF-----RVPYVMSATCSCGFLGLVYCTPGGQYILE 457  
DB 389 CLGIGSOVYGVGYTALIDLPMSFESFRYLRVYVAVAVICLSFLIDLSMTTEGMYIFQ 448  
QY 458 LVNDY--GGTFLVLECAISELAGVMIYGLNCLDIEFPLGKKTGAWRICMGVITPAI 515  
DB 449 IFDYVSSAGTTL-LMTAFMECVAVAMVYGGDRYIDDLAMNLGYRPMALVYKRCWSTITPLV 507

QY 516 MTTVFYFALLASNVLFGDNVYPTAGVSGYLMFLGMTPFVPIGIFSLKRYRGTSE 575  
DB 508 CMGIFLFLVYKPLTYNNTYTYPPWGEAIGCLALASMLCVPTVYLSLSGR-GSLKE 566  
QY 576 TIKKAHSPKSPGPRSPREREMQFKAQKALRQKMNSRYKHLWYSTGA 627  
DB 567 RMRKTL-TTPVWASH-----HLAYKAGA 588

## RESULT 9

F46027  
gamma-aminobutyric acid transporter - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Feb-1997  
C:Accession: F46027  
R:Liu, Q.R.; Mandiyan, S.; Nelson, H.; Nelson, N.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6639-6643, 1992  
A:Title: A family of genes encoding neurotransmitter transporters.  
A:Reference number: A46027; MUID:92335351; PMID:1631167  
A:Accession: F46027  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-598 <LIU>  
A:Experimental source: brain  
A:Note: sequence extracted from NCBI backbone (NCBIN:112460, NCBIPI:108818)  
C:Superfamily: gamma-aminobutyric acid transporter  
C:Keywords: transmembrane protein

## Query Match

29.2%; Score 978.5; DB 2; Length 598;  
Best Local Similarity 34.3%; Pred. No. 2.6e-67;  
Matches 208; Conservative 123; Mismatches 232; Indels 43; Gaps 13;

QY 3 DGQVNGGSESPKMKPRSSQIS-LPPANAKALDNIDDDLEAEPERWMSNIEFL 61  
DB 10 DGOI-----STEVSEAPVASDKPTLYVVKQKADLDPRDPTWKGRPFL 54  
QY 62 MSCIATSVGLGNWREFFIAYONGGAFIVPYVILLVGRPYVLECVLOFSSRSYK 121  
DB 55 MSCVYATGLGNWREFFIAYONGGAFIVPYVILLVGRPYVLECVLOFSSRSYK 114  
QY 122 WMSISPAKMGTYAQAAGCYIIISYYVIGCLLYLAMSPQATLPAIICQPEW--ENCY 179  
DB 115 VWNMAIPMKGVAVAAVLSFPMNIYIYIISAIYLVKNSFTTILPMKQCONPMNTDCE 174  
QY 180 PSDPTLASVNNITNGTSSAQIYFLKTVLQOSDGIIEGIGAPIWTVLYCLFIAMLMVEGV 239  
DB 175 SN-----YSLVNTTMTSAVVEFMERNHOMTDGID--KPGQIRCLAITLIAIAYLYFC 227  
QY 240 VARGVSSGKAAYFLALFPYVMTLFTTTIILPGADGILFPYVPOKAKILEGVMSA 299  
DB 228 IMKGVGMKGVYFSAIYPIYIMLITLFRGVTLPQAKGILFTYTPNRKLSDSSEVTPDA 287  
QY 300 VTQVFFSLVCTGPIIMSSYNGFRHNIRDAWITTLDTPTSLSGCTIFIGLNLAYE 359  
DB 288 ATQLEFFSYGLGIGSLIASYNSFHNHYRDSIIYCCINCSQSMFAGVIFSIYGFMAHV 347  
QY 360 LNSEVDVVGAGTSLAFISYPAIAKTFOPOLSVLEFLMMSVYIGISSVALLSTFNTL 419  
DB 348 TKRSIAD-VAASGPGIAFIAPKAVITLMPFQVAVLEFIMLLMIGIDSDCTVEGFTTA 406  
QY 420 AMDAPFPRV-----PYVMSAMTSCSGFLGLVYCTPGGQYILEVDHYGGTFL-VLFCALS 474  
DB 407 LVDEYPRLLRRRRELFIAAV-CIYSYLGSLNITGQYIYKRLDYYSASMSLFLVLF 465  
QY 475 ELAGVFTYIGLENCLDIEFPLGKKTGAWRICMGVITPAIMTTFEYFALLASNVLFGD 534  
DB 466 ECVSISWYGVNRPFDNIQENAVGSRPCIMWKILCSFPIIYVAGVIFLSAVQMPPLMG- 524  
QY 535 NYVYPTAGVSGYLMFLGMTPFVPIGIFSLKRYRGTSEIYKRT--GTSEIYKKAHFS-----RPSW 587  
DB 525 SYVPPKMGQGVIMLALSSVNLIP---GYMAVYFLLTKGSLKQRLQVNIQPSSEDIYVRPN 581

Oy 588 GPRSPR 593  
 Db 582 GPEPO 587

## RESULT 10

gamma-aminobutyric acid transporter protein 3 - rat  
 JH0695  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 20-Aug-1999  
 C:Accession: JH0695; B45078  
 R:Clark, J.A.; Deutch, A.Y.; Gallipoli, P.Z.; Amara, S.G.  
 Neuron 9, 337-348, 1992  
 A:Title: Functional expression and CNS distribution of a beta-alanine-sensitive neuronal  
 A:Reference number: JH0695; MUID:92360310; PMID:1497897  
 A:Accession: JH0695  
 A:Molecule type: mRNA  
 A:Residues: 1-627 <CLA>  
 A:Cross-references: GB:S42358; NID:9253466; PIDN:AA2850.1; PID:9253467  
 R:Biogen, L.A.; Smith, K.E.; Hartig, P.R.; Branchek, T.A.; Weinstank, R.L.  
 J. Biol. Chem. 267, 21098-21104, 1992  
 A:Title: Molecular heterogeneity of the gamma-aminobutyric acid (GABA) transport system.  
 A:Reference number: A45078; MUID:93016029; PMID:1400419  
 A:Accession: B45078  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-627 <BOR>  
 A:Cross-references: GB:M95763; NID:9202534; PIDN:AAA0607.1; PID:9202535  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBI:P116200)  
 C:Superfamily: gamma-aminobutyric acid transporter  
 C:Keywords: brain; glycoprotein; phosphoprotein; transmembrane protein  
 F:54-74/Domain: transmembrane #status predicted <TM1>  
 F:81-101/Domain: transmembrane #status predicted <TM2>  
 F:126-146/Domain: transmembrane #status predicted <TM3>  
 F:223-239/Domain: transmembrane #status predicted <TM4>  
 F:248-270/Domain: transmembrane #status predicted <TM5>  
 F:297-317/Domain: transmembrane #status predicted <TM6>  
 F:330-351/Domain: transmembrane #status predicted <TM7>  
 F:383-403/Domain: transmembrane #status predicted <TM8>  
 F:433-451/Domain: transmembrane #status predicted <TM9>  
 F:468-488/Domain: transmembrane #status predicted <TM10>  
 F:509-528/Domain: transmembrane #status predicted <TM11>  
 F:548-566/Domain: transmembrane #status predicted <TM12>  
 F:182,185,193/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:505,598/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:603/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 29.0%; Score 973; DB 2; Length 627;  
 Best Local Similarity 35.8%; Pred. No. 7.2e-67;  
 Matches 209; Conservative 109; Mismatches 236; Indels 30; Gaps 11;  
 Oy 22 SSQSLPPANKKALDNDITDLDL-----EAEP-----ERWVNNNIEFLMSCIAT 67  
 Db 2 TAEQALPLPNGKAEEARGSEALGGGGAAGTRARDKAVHERHNNKKEFLVSLVAGE 61  
 Oy 68 SVGLGAVWRPFPIAIONGGAFVLYVLLVGVYLLVLCVGOFSRRSVKWW-SIS 126  
 Db 62 IITGLNWRPFPLCKKNGGAFILPYVFFICGIPVFFLETALGOFSEGGITCKRVC 121  
 Oy 127 PAMKGTGAQAAGCYLLSYVYVIGLCLVYLLANSFOATLPWALICQPEW--ENCVPSPDT 184  
 Db 122 PLEEGIGATQVIEHNLVNYIILAMAFYLSNCFTELPWATCGHEMNEKCYEFOKL 181  
 Oy 185 LAASVNNIT--NGTSSADLYFLRYLQSDGIEGGLAPITVLYLCEFIANLWVGVAR 242  
 Db 182 NFNNSHVSLOANAPSVPMEFERVLAISDGE-HIGMLRELACILAAWTICFCIKW 240  
 Oy 243 GVKSGKAAYFLALFPYVYVITFTTIIIPGATDGLFVTPQAKLLLEGWVASVTO 302  
 Db 241 GTSKSGKVVYVATFPYITMLILIRGVTLFQASBGIFLYLPDLSTRISLQVWVDAQTO 300

Oy 303 VFESLIVCTGPIIMFESSYNGFRHNIYRDAMLVTTLDITFSLSCGTFIGLGNLAYELNS 362  
 Db 301 IFESYALCLGCLRALGSIYNNYNNNCYRDCIMLCCLNGSTSFVAGFAIFSLGPMAYQGV 360  
 Oy 363 EVGDVAGAGTSLAFISYPDAIAKTFQPOLESVLFPLMSVLAGISSVALLSTFNTLMD 422  
 Db 361 PIAE-VAESGGLAFIAYPRKAVTMTPLMTLFFMLIFGLDGSQFCVCSLVFAVVD 419  
 Oy 423 AFPRV-----PIYVMAAMCSCGFLGLVYCTFGGOYILEVDHNGT-FLVIFCAISE 475  
 Db 420 MYPRFRRGYRRRELLIALSL-VSYFLGLVNLTEGMYIFQLFDSVAAAGCLLEVAIFE 478  
 Oy 476 LAGVFVYIGLENLCLDIEFMIGKRTGAYWRLCWGVIPATITVFFYFALLASNNLVFGDN 535  
 Db 479 CVCIGWYYSNRFPDNIEMDGMGRPLSLIKMKWVYVPGICAGIFIEFVLYKRLKNNV 538  
 Oy 536 VYPTAGYVSGYLMFLGMPFVPIGIFSLLYKRTGFESETIKK 579  
 Db 539 YTYPMAGYIGIGMLMALSSMLCTIPMIFIKIKW-TEGTLPERLQK 581

## RESULT 11

choline transporter protein - rat  
 S23431  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 20-Aug-1999  
 A:Accession: S23431  
 R:Mayser, W.; Schiöss, P.; Betz, H.  
 FEBS Lett. 305, 31-36, 1992  
 A:Title: Primary structure and functional expression of a choline transporter express  
 A:Reference number: S23431; MUID:92339519; PMID:1633856  
 A:Accession: S23431  
 A:Molecule type: mRNA  
 A:Residues: 1-635 <MYV>  
 A:Cross-references: EMBL:X66494; NID:955948; PIDN:CAA47119.1; PID:955949  
 C:Superfamily: gamma-aminobutyric acid transporter  
 C:Keywords: glycoprotein; transmembrane protein  
 F:61-83/Domain: transmembrane #status predicted <TM01>  
 F:88-108/Domain: transmembrane #status predicted <TM02>  
 F:133-160/Domain: transmembrane #status predicted <TM03>  
 F:233-251/Domain: transmembrane #status predicted <TM04>  
 F:260-277/Domain: transmembrane #status predicted <TM05>  
 F:313-330/Domain: transmembrane #status predicted <TM06>  
 F:342-366/Domain: transmembrane #status predicted <TM07>  
 F:396-415/Domain: transmembrane #status predicted <TM08>  
 F:445-462/Domain: transmembrane #status predicted <TM09>  
 F:483-500/Domain: transmembrane #status predicted <TM10>  
 F:521-542/Domain: transmembrane #status predicted <TM11>  
 F:560-581/Domain: transmembrane #status predicted <TM12>  
 F:192,197,548/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 29.0%; Score 972; DB 2; Length 635;  
 Best Local Similarity 35.8%; Pred. No. 8.7e-67;  
 Matches 209; Conservative 103; Mismatches 228; Indels 44; Gaps 11;  
 Oy 22 SSQSLPPANKKALDNDITDLEAEPERWVNNNIEFLMSCIATSYGLGNVRRPPIA 81  
 Db 44 SSRALVPP-----RETWRQMDPIMSCVGFAGLGNVRRPFLYC 82  
 Oy 82 YONGGAFVLPYVYVLLVGVYLLVLCVGOFSRRSVKWWSPAKKGTGAQAAGCG 141  
 Db 83 YKNGGVLPVLYVLLVAGGIPFFLESLSDQFPMKAGSINWNCIPLEKGLGVASWYVF 142  
 Oy 142 YLSYVYVIGLCLVYLLAMSFOATLPWALICQPEW--NCVP---SDPLTAAVNNIT--- 193  
 Db 143 YCNITTYINVLAMGFYLYLKSFTTILPMAICGHTNNTPPCVLIFRHEDCANNSLANLTCQD 202  
 Oy 194 --NGTSSADLYFLRYLQSDGIEGGLAPITVLYLCEFIANLWVGVARGVKSGKAA 251  
 Db 203 LADRRSPVIEFEMENKVLRLSTGLEVP-CALNMEVITCLACQVILVYFCVWKGKSTKIV 261  
 Oy 252 YFLALFPYVYVITFTTIIIPGATDGLFVTPQAKLLLEGWVASVTOVFPFSLVCT 311

```
Db      262 YFTAFYFFVLLVVLVRGVLLPGALDGIYYLKPDMSKIGSPQWIMDACTQIFFYSAICL 321
Oy      312 GPIIMFESSNGFRHHIYRDAMIIVTTLDTFTSELSCGTIFGILGNLAIELNSEVDVAG 371
Db      322 GALTALGSYNRFNNNCYKDAILIALINSGTSPEFAGVFVSILGEFMATEGVHISK -VAES 380
Oy      372 GTSLAFTISPRAIATFPQPLFSVFLPMLMASVIGIGSSALLSTFNTLAMADP-----R 426
Db      381 GPGLAEIYFPRAVITLMPAPVAPMAALEEFEMILLGLDSOPVEGEGFTTGIDLLPASYYFR 440
Oy      427 VPITYMASMTSCSCEFLGLSVYTCPGGVYLLELVDRH--GCFELVLCASIELAGFWING 484
Db      441 FQRETSVALCCALCVCVIDLSMTDBGMVYFQLEDYISAGTTL-LMQARWECVVANMYIG 499
Oy      485 LENICLDIEFMLGKRTGAIVRLCWGVITPAINTVFYFALLASNULVPFGDNVYPTAGVY 544
Db      500 ADREFWDJACMIAGYRPCPMWKMCSEFFPLYOCMGEIFFNVVYXXKLPLYNNTYYPPMWGEA 559
Oy      545 SGYLMFLQMTPFVPIGIGFSLKYKRGTFSEIKRAHFHKPSMG 588
Db      560 MGMFAFALSMLCVPHLLGLCLRAR-GVIAEWMQH.-NQPLWG 600
```

## RESULT 12

serotonin transport protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C:Accession: S19585  
R:Blakely, R.D.; Berson, H.R.; Freneau Jr., R.T.; Caron, M.G.; Peek, M.M.; Prince, H.K.  
Nature 354, 66-70, 1991  
A:Title: Cloning and expression of a functional serotonin transporter from rat brain.  
A:Reference number: S19585; MUID:92045754; PMID:1944572  
A:Accession: S19585  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-607 <BLA>  
A:Cross-references: EMBL:X63253  
C:Superfamily: gamma-aminobutyric acid transporter  
;Keywords: transmembrane protein

[illegible]

```

QY      430 YYMSAMTSCSGRLGLGVYTCPCGGQYILELVHDYHGGLVLFCAISLACVFWIYGLENC 489
Db      443 VLIIVITCVLGSLLT---TSGAGVYVTLLEEATGPAVLVALLIAVAVSMWYGIQFC 499
QY      490 LDIEMLGKRTGAVYWRCLGVTITPAINTVFYFALLASNNL-VEGNNYVPTAGYSGYL 548
Db      500 SDVKEMLEGPSPGFMRICWVAISPFLFIISFIMSPPOLRIE-QYNNPHMSIVLGYC 557
QY      549 MLFLMTFVPIGIGISLVYKRTGTSETIKKAFAHSPKSPGMRPRE 594
Db      558 IGMSSVICPIYIIRKLIS-TPGTLKERIK-----STTPETPT 596

```

## RESULT 13

creative transporter - human  
 N:Alternate names: solute carrier family 6, member 8  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
 C:Accession: G02095  
 R:Elchler, E.E.; Lu, F.; Muzny, D.; Gibbs, R.A.; Nelson, D.L.  
 submitted to the EMBL Data Library, September 1995  
 A:Reference number: G09194  
 A:Accession: G02095  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-635 <EID>  
 A:Cross-references: EMBL:U36341; NID:g1020318; PIDN:AAA79507.1; PID:g1020319  
 C:Genetics:  
 A:Gene: GDB:SLC6A8; CF-1  
 A:Cross-references: GDB:434311; OMIM:600394  
 A:Map position: Xq28-Xq28  
 A:Introns: 88/1; 132/1; 215/2; 259/3; 304/3; 339/2; 381/1; 418/3; 464/3; 499/1; 532/3  
 C:Superfamily: gamma-aminobutyric acid transporter

Query Match	29.0%	Score 971.5	DB 2	Length 635
Local Similarity	37.0%	Pred. No. 9,5e-67		
Matches	206	Conservative 99	Mismatches 229	Indels 23
			Gaps 10	
QY	PERMWSNNIEFASCIATSVGLGWNMFPEFLAYONGGAFVLPVIVLLVGRPVYLE	108		
DB	50 PPREWTVQDMFINSCHVFAVLGNWVWFPLICYNGSGVLELIPVLLALVGGIPFLE	109		
QY	109 CVLGQSSRSNYSKWSISIPAMKGTGYAQAAGCGYLLSYVVYICGLCLYLAASFQATLPW	168		
DB	110 ISLQGFMAKGSINWNICPLFRKGIGYASAMVYFQNTYTYIMVLAMGEFYLVKSFPTLPM	169		
QY	169 AICQPEWF--NCVP---SDPTLAASVNNIT-----NGTSAQLYFLRTVLQSDGIEGGL	218		
DB	170 ATCGHTWTVPCDCEIFRHEDCANASLMTCDQLADRRSPVIEFENKVLRLRISGGLLEVP-	228		
QY	219 GAPIMVYLVLCFIAMLMFVGVYARVYKSSGRAAYFLALFPVYMTLTFITILLPGALDG	278		
DB	229 GALNNEVLTCLLACVNLVYFCVWVKGKSTGKIYFTAFPPVYVLLVLRGLVLPALDG	288		
QY	279 ILFETPOMAKLLELVNYSAVTQVQFSLYCTQGPINMESNGFRHNIYDAMVTITLD	338		
DB	289 IITYIKRPMWSKLSQVWIDAGTQTFEFSYALIGLALTLAGSYNFRNNCKIDALLALIN	348		
QY	339 TETSELSCGCTIFGILGNLAYELNSEHVGGVNAGGATSLAFISYDPAIARTFQPOLESVLEF	398		
DB	349 SGTSEFAGFVYFSLIGFMAAEGVHISK-VAESGGGLAFIAYPRAVTILMPAPLMAALEF	407		
QY	399 LMSVYLIGSSVALLISTNTLAMDAF-----RPVTYMSAMTSCGFLGLGIYCTPGQ	453		
DB	408 FMILLIGLGDSPVGEVEGFTGLDLDPASYYFRFQREISVALCCALCEVIDLSMTDCGM	467		
QY	454 YIIELVNDY--GGFTVLVFCALISELAGFVNYIGLENCLDIEFMLGKGTGAYMRLCNQVI	511		
DB	468 YVFQLEFDYYSASGTTL-LMQAFWECCVVAMVYAGAPRFMDIACMGYPRCPMKNCSFF	526		
QY	512 TPAINTVYVFALLASNNLVFGDNVYVPTAGVYSGYLMFLQMTFVPIGIGESFLYKYRTG	571		



OY 572 TFSSTIKKAFHSKPSWG 588  
| : | : : | |  
Db 586 TMAERWQHL-TOPTWG 600

Search completed: July 21, 2003, 09:26:11  
Job time : 45 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:24:45 ; Search time 24 Seconds

(without alignments) 1093.938 Million cell updates/sec

Title: US-09-991-458-2  
Perfect score: 3354  
Sequence: 1 MNDQVNGGESESEPKMEPK.....TSRVKHLWYSITGAYRRNIN 633

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1080.5	32.2	797	1	Q9345_HUMAN
2	1076.5	32.1	799	1	Q9345_HUMAN
3	1027	30.6	637	1	Q9345_HUMAN
4	1020	30.4	636	1	Q9345_HUMAN
5	1011	30.1	599	1	Q9345_HUMAN
6	1008	30.1	599	1	Q9345_HUMAN
7	998	29.8	630	1	Q9345_HUMAN
8	992.5	29.6	598	1	Q9345_HUMAN
9	991	29.5	630	1	Q9345_HUMAN
10	991	29.5	630	1	Q9345_HUMAN
11	987	29.4	630	1	Q9345_HUMAN
12	981	29.2	630	1	Q9345_HUMAN
13	979	29.2	614	1	Q9345_HUMAN
14	978.5	29.2	598	1	Q9345_HUMAN
15	978.5	29.2	632	1	Q9345_HUMAN
16	975	29.1	635	1	Q9345_HUMAN
17	974.5	29.1	620	1	Q9345_HUMAN
18	973	29.0	627	1	Q9345_HUMAN
19	972	29.0	635	1	Q9345_HUMAN
20	971.5	29.0	635	1	Q9345_HUMAN
21	971	29.0	622	1	Q9345_HUMAN
22	970	28.9	630	1	Q9345_HUMAN
23	968.5	28.9	630	1	Q9345_HUMAN
24	967	28.8	620	1	Q9345_HUMAN
25	965.5	28.8	620	1	Q9345_HUMAN
26	965	28.8	602	1	Q9345_HUMAN
27	964	28.7	614	1	Q9345_HUMAN
28	963	28.7	627	1	Q9345_HUMAN
29	959	28.6	602	1	Q9345_HUMAN
30	953.5	28.4	621	1	Q9345_HUMAN
31	949.5	28.3	620	1	Q9345_HUMAN
32	949.5	28.3	635	1	Q9345_HUMAN
33	943.5	28.1	621	1	Q9345_HUMAN

34	940.5	28.0	614	1	S6A3_MOUSE
35	939	28.0	619	1	S6A3_MOUSE
36	936	27.9	615	1	S6A3_MOUSE
37	936	27.9	638	1	S6A3_MOUSE
38	935	27.9	633	1	S6A3_MOUSE
39	935	27.9	633	1	S6A3_MOUSE
40	934.5	27.9	621	1	S6A3_MOUSE
41	933	27.8	619	1	S6A3_MOUSE
42	932.5	27.8	614	1	S6A3_MOUSE
43	931.5	27.8	614	1	S6A3_MOUSE
44	930	27.7	617	1	S6A3_MOUSE
45	920.5	27.4	617	1	S6A3_MOUSE

## ALIGNMENTS

RESULT 1  
ID S6A5\_HUMAN STANDARD: PRT: 797 AA.  
AC Q9345: 095288; Q9BX77;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Sodium- and chloride-dependent glycine transporter 2 (GLYT2) (GLYT-2).  
GN SLC6A5 OR GLYT2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99059482; PubMed=9845349;  
RA Morrow J.A., Collie I.T., Dundar D.R., Walker G.B., Shahid M.,  
RA Hill D.R.;  
RT "Molecular cloning and functional expression of the human glycine  
RT transporter GLYT2 and chromosomal localisation of the gene in the  
RT human genome.";  
RL FEBS Lett. 439:334-340(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99310916; PubMed=10381548;  
RA Gallagher M.J., Burgess L.R., Brunden K.R.;  
RT "Characterization of multiple forms of the human glycine transporter  
RT type 2.";  
RL Brain Res. Mol. Brain Res. 70:101-115(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Luyten W.;  
RT "Cloning and expression of a human glycine transporter type II.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A., AND VARIANTS ARG-184; ASN-463 AND ALA-751.  
RX MEDLINE=20074920; PubMed=10606742;  
RA Evans J., Herdon H., Cairns W., O'Brien E., Chapman C., Terrett J.,  
RT "Cloning, functional characterisation and population analysis of a  
RT variant form of the human glycine type 2 transporter.";  
RL FEBS Lett. 463:301-306(1999).  
CC -1- FUNCTION: Terminates the action of glycine by its high affinity  
CC sodium-dependent reuptake into presynaptic terminals. May be  
CC responsible for the termination of neurotransmission at  
CC strychnine-sensitive glycinergic synapses.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC in spinal cord and cerebellum.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER  
CC FAMILY (SNP).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF085412; AAC95145.1; -  
 DR EMBL: AF142501; AAD27892.1; -  
 DR EMBL: AF117999; AAK12641.1; -  
 DR EMBL: AF352733; AAK29670.1; -  
 DR Genew: HGNC:11051; SLC6A5.  
 DR MIM: 604159; -  
 DR InterPro: IPR000175; Na/nttran\_symport.  
 DR Pfam: PF00209; SNF; 1.  
 DR ProDom: PD000448; Na/nttran\_symport; 2.  
 DR PROSITE: PS00610; NA\_NEUROTAN\_SYM\_1; 1.  
 DR PROSITE: PS00734; NA\_NEUROTAN\_SYM\_2; 1.  
 DR PROSITE: PS50267; NA\_NEUROTAN\_SYM\_3; 1.  
 DR Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
 KW Symport; Polymorphism.

FT DOMAIN 1 199  
 FT TRANSMEM 200 220 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 228 247 1 (POTENTIAL).  
 FT TRANSMEM 271 291 2 (POTENTIAL).  
 FT DOMAIN 292 393 3 (POTENTIAL).  
 FT TRANSMEM 394 412 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 421 438 4 (POTENTIAL).  
 FT TRANSMEM 474 491 5 (POTENTIAL).  
 FT TRANSMEM 503 524 6 (POTENTIAL).  
 FT TRANSMEM 557 576 7 (POTENTIAL).  
 FT TRANSMEM 604 622 8 (POTENTIAL).  
 FT TRANSMEM 638 658 9 (POTENTIAL).  
 FT TRANSMEM 679 698 10 (POTENTIAL).  
 FT TRANSMEM 717 735 11 (POTENTIAL).  
 FT TRANSMEM 737 797 12 (POTENTIAL).  
 FT CARBOHYD 343 343 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 184 184 O -> R.  
 FT VARIANT 463 463 FTID-VAR\_011591.  
 FT VARIANT 751 751 D -> N (IN 10% OF THE POPULATION).  
 FT VARIANT 751 751 FTID-VAR\_011592.  
 FT VARIANT 751 751 V -> A.  
 FT CONFLICT 24 24 FTID-VAR\_011593.  
 FT CONFLICT 102 102 G -> S (IN REF. 2).  
 FT CONFLICT 155 155 G -> S (IN REF. 2).  
 FT CONFLICT 188 188 S -> G (IN REF. 2).  
 FT CONFLICT 362 362 N -> D (IN REF. 2).  
 FT CONFLICT 582 582 Q -> L (IN REF. 2).  
 FT CONFLICT 582 582 T -> S (IN REF. 2).  
 SQ SEQUENCE 797 AA; 87359 MW; BBA089AB7EBBDA CRC64;

Query Match 32.28; Score 1080.5; DB 1; Length 797;  
 Best Local Similarity 36.28; Pred. No. 7.2e-64;  
 Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10;

QY 54 MSNIEFLMSCIATSVGIGNWRPPTAYONGGAGFAPYIVLLVGRPYVLECYGQ 113  
 DB 194 WSKSLDFILSVNGAVGLGNWRPPTAYONGGAGFAPYIVLLVGRPYVLECYGQ 253  
 QY 114 FSSNSVYKWSISPMKGTGYAAGCGYILSYVVICGLCLVYLLANSFOATLFWAICOP 173  
 DB 254 FASGQPSVWKAIPALOGCGIAMLIIISVLAIYVNVIICTLFYLFASFSVLPWGCNN 313  
 QY 174 EW-----ENCVSD-PTL-----AASVNNIT-NGISSA-----QL 201  
 DB 314 PMNTECKDKTKLLLDSCVISHDKPIQIKSTFCMTAYPNVTWNFTSQANKTFVSGSEE 373  
 QY 202 YFLRTVLOQSDIEGGLGAPITWVLCLEFAMLVFEGVANGVSSGKAAYFLALFPYV 261  
 DB 374 YFKFVLIKISAGIEYR-GEIRWPLALCLFLAMVIVYASLAKGITSKVVYITFTFFYV 432

QY 262 MIFLFTITLLPGATDGIIFPVTPONAKLLELVWYSAAVOFEISLVTCTPIFMSSYN 321  
 DB 433 LVILLRGVLLPAGAGACITWFTPKMKELTDAYMDAAQIFFSLSAMGGLITLSSYN 492  
 QY 322 GFRHNIYRDAMIYTTLDFTSFLSGCTIFGILGNLYELNSVGDVAGAGTSIATISP 361  
 DB 493 KFRHNCYRDOLIYCTNSATSIIFAGVIFSVIGFMANERKVIENVADQ-GPGIARVYVP 551  
 QY 382 DATAKTPQPLESVLPFMSVIGISGVALSTFNPLANDAPRPVYMSMT---CS 438  
 DB 552 EALTRLPLSPFWMIIFFLMLTLGLDMEFTIETVTSISDEPKYLRTHKPVFTLACCT 611  
 QY 439 CGFLGIVYPTCGOVYLLVDHYGGTFLVFCALISLACVFNLYGLENLCLDIEMPLGK 498  
 DB 612 CFPIMGPMITOGGIWYFQVLDVRYAASVALYIAIFELVGISVYGLQRCEDEIMWIGF 671  
 QY 499 KTGAYNRLCGVITPAIMTTFVYFALLASNNLVFGDNVYPTGIYSGYIMLEFGMTVP 558  
 DB 672 QPNIFMKVCAWVYPTLITFLICFSFYQWPMYTG-SYRYPNMSMYLGLMLACSVIWP 730  
 QY 559 IGIGFSLYKRYTGFESETIKAFHNSKPSWGRSPRERRE 597  
 DB 731 IMFYIKMH-LAPGRFIERLKVCSPPDWPGLAQRGE 768

RESULT 2  
 ID 56A5\_RAT STANDARD; PRT; 799 AA.  
 AC P58295;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sodium- and chloride-dependent glycine transporter 2 (GlyT2) (GlyT-2).  
 GN SLC6A5 OR GLYT2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94043045; PubMed=8226790.  
 RA Liu Q.-R., Lopez-Corcuera B., Mandlyan S., Nelson H., Nelson N.;  
 RT "Cloning and expression of a spinal cord- and brain-specific glycine  
 RT transporter with novel structural features.";  
 RL J. Biol. Chem. 268:22802-22808(1993).  
 CC -|- FUNCTION: Terminates the action of glycine by its high affinity  
 CC sodium-dependent reuptake into presynaptic terminals. May be  
 CC responsible for the termination of neurotransmission at  
 CC strychnine-sensitive glycinergic synapses.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -|- TISSUE SPECIFICITY: Specifically expressed in spinal cord, brain  
 CC stem, and to a lesser extent in the cerebellum.  
 CC -|- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER  
 CC FAMILY (SNF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L21672; -; NOT\_ANNOTATED\_CDS.  
 DR InterPro: IPR000175; Na/nttran\_symport.  
 DR Pfam: PF00209; SNF; 1.  
 DR PRINTS: PR00176; NANEUSPORT.  
 DR PROSITE: PS00610; NA\_NEUROTAN\_SYM\_1; 1.  
 DR PROSITE: PS00734; NA\_NEUROTAN\_SYM\_2; 1.  
 DR PROSITE: PS50267; NA\_NEUROTAN\_SYM\_3; 1.  
 KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
 KW Symport.

```

FT DOMAIN 1 201 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 202 222 1 (POTENTIAL).
FT TRANSMEM 230 249 2 (POTENTIAL).
FT TRANSMEM 273 293 3 (POTENTIAL).
FT DOMAIN 294 395 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 396 414 4 (POTENTIAL).
FT TRANSMEM 423 440 5 (POTENTIAL).
FT TRANSMEM 476 493 6 (POTENTIAL).
FT TRANSMEM 505 526 7 (POTENTIAL).
FT TRANSMEM 559 578 8 (POTENTIAL).
FT TRANSMEM 606 624 9 (POTENTIAL).
FT TRANSMEM 640 660 10 (POTENTIAL).
FT TRANSMEM 681 700 11 (POTENTIAL).
FT TRANSMEM 719 737 12 (POTENTIAL).
FT DOMAIN 739 799 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 799 AA: 87908 MW: 9DBE239A7B071FCD CRC64:

```

Query Match 32.1%; Score 1076.5; DB 1; Length 799;  
 Best Local Similarity 37.5%; Pred. No. 1.3e-63;  
 Matches 224; Conservative 107; Mismatches 228; Indels 39; Gaps 10;

```

QY 35 ALDNIDTDLAEPEPERWWSNNIEFLMSCIATSVGLGNWFFPIAYONGGAFLEVPYV 94
DB 177 ATTITPEDEQDENAKRGMSKDLFIISMVGYANGLGWVFPIATONGGAFLEVPYV 236
QY 95 IVLLVGRPVYILECVLQFSSRSNYSVSPANKGTGYAQAAGCYILSYVYVIGCIG 154
DB 237 MMLALAGLPFEFLESLGQFASQGPVSWKAIPALQCGIAMIILSVLAIVYNYICTY 296
QY 155 LYVILAMSPQATLPMAIQCPW-----ENCVPSD-PTL-----ASVNNI 192
DB 297 LEYLFASFVSVLPWQSCNNPMPTPECKDKTKLLDSVYGDHPKIQINKSTFCMTAYPM 356
QY 193 T--NCTSSA-----OLYFLRTVLOOSDGIIEGLGAPITWLVLCLFAMLMVEGVAR 242
DB 357 TMVNTSQANKTFVSGSEYEFYFKISAGIEYF-GELRWLPPLCFLFAMVIVYASLAK 415
QY 243 GYKSSGKAAVFLAFPPYVMTLFTITLLPGATDGLLEFVTPQNAKLLLELGVWSAVTQ 302
DB 416 GIKTSGKVVYFATPPYVVLVILIRGYLLPGAGAGIWFETPKWEKLLDATVWMDAATQ 475
QY 303 VFEFLVTCPTIMSSVNGFRHNYRDAMVITLDTFTLSGCTIRIILNLNAYELS 362
DB 476 IFESLSAAMGGLITLSTINKFRNNCYRDLVITGNSATSIAGFVIESVIGFMANERKV 535
QY 363 EYGDVYVAGGTSIAFISYPDAIAKTFQDLPFSLFELMMSVGLSSVALLSTFNTLIAM 422
DB 536 NIENNVADQ-GPGIATVYVPEALTRPLSPFMAIIFEMILLTGIDLMFTIETIYISID 594
QY 423 AFRPVYVYMSAMT---CSCGFLGLVYCTPGGQYILLVHDYGGTFVLVFCALISLAV 479
DB 595 EEPKYLIRHKPVFTLGGCCICEFIMGFPMTQGGIMFOLVDVYASAYALVIAIFELVGI 654
QY 460 FVIYGLNLCIDIEMLGKKGAVWRICGVITPAMITVFFYALLASNNLVGQNYVYP 539
DB 655 STVYGLQRCDEIEMIGFQPNIFMKVCAVFTPTLLIFLCEFSFYQWEPMTYG-SYRP 713
QY 540 TAGVYGLMFLFTGTFVPIGIGFLYKRTGTFTSETIKAKHSPMSGPRPRERE 597
DB 714 NMSVNLGMLMACSYIWMIPIMVIMKY-LAPGRFIERLKLVCSPQDMDGPFLLAQHGE 770

```

RESULT 3  
 56A7\_RAT  
 ID 56A7\_RAT STANDARD: PRT: 637 AA.  
 AC P28573:  
 DT 01-DEC-1992 (rel. 24, Created)  
 DT 16-OCT-2001 (rel. 40, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)

```

DE Sodium-dependent proline transporter.
GN SLC6A7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92265305; PubMed=1350201;
RA Fremieu R.T., Jr., Caron M.G., Blakey R.D.;
RT "Molecular cloning and expression of a high affinity L-proline
RT transporter expressed in putative glutamatergic pathways of rat
RT brain."
RL Neuron 8:915-926(1992).
CC -1 FUNCTION: TERMINATES THE ACTION OF PROLINE BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -1 SUBCELLULAR LOCATION: Integral membrane protein.
CC -1 TISSUE SPECIFICITY: EXPRESSED IN SUBPOPULATIONS OF PUTATIVE
CC GLUTAMATERGIC PATHWAYS OF RAT BRAIN.
CC -1 SIMILARITY: BELONGS TO THE SODIUM-NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNP).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M8811; AAA41541.1; ALT TERM.
CC Interpro: IPRO00175; Na/ntan_symport.
CC Pfam: PF00209; SNF. 1.
CC PRINTS: PRO0176; NANEUSMPORT.
CC PRODOM: PD000448; Na/ntan_symport; 1.
CC PROSITE: PS00610; NA_NEUROTRAN_SYM_1; 1.
CC PROSITE: PS00754; NA_NEUROTRAN_SYM_2; 1.
CC PROSITE: PS50267; NA_NEUROTRAN_SYM_3; 1.
CC KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
CC Symport; Amino-acid transport.
CC -----
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT TRANSMEM 74 93 2 (POTENTIAL).
FT TRANSMEM 117 137 3 (POTENTIAL).
FT DOMAIN 138 214 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 215 223 4 (POTENTIAL).
FT TRANSMEM 242 259 5 (POTENTIAL).
FT TRANSMEM 295 312 6 (POTENTIAL).
FT TRANSMEM 324 345 7 (POTENTIAL).
FT TRANSMEM 378 397 8 (POTENTIAL).
FT TRANSMEM 425 443 9 (POTENTIAL).
FT TRANSMEM 455 479 10 (POTENTIAL).
FT TRANSMEM 500 519 11 (POTENTIAL).
FT TRANSMEM 538 556 12 (POTENTIAL).
FT DOMAIN 557 637 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 637 AA: 71090 MW: 9627E6D5B8C9408 CRC64:

```

Query Match 30.6%; Score 1027; DB 1; Length 637;  
 Best Local Similarity 39.8%; Pred. No. 1.9e-60;  
 Matches 227; Conservative 90; Mismatches 236; Indels 18; Gaps 11;

```

QY 41 DDDLEAF-PPERWWSNNIEFLMSCIATSVGLGNWFFPIAYONGGAFLEVPYVILL 99
DB 26 DVDLDVDFADAGNMTGKLDLPFLSCIGYVGLGNWVFPYRAYTNGGAFLEVPYVILL 85
QY 100 VGRPVYILECVLQFSSRSNYSVSPANKGTGYAQAAGCYILSYVYVIGCIGLYYA 159
DB 86 CGPLPFELESLGQFSSLSGPPLAVWKSPLFKGAGAMLLVGLVAYVMMIAYVLFYLF 145
QY 160 MSFQATLPMAIQCPW-----PSPDTLAASVNNITNGTSSAQLYFLRTVL--QQS 211

```

Db 146 ASLTSNLPWEHCGMMWNTERCLERHPKNGALPL-NLSTVSPSEBYRSRYVLIHIGS 204  
QY 212 DGIIGGLGAPIMYLVLCFLIAMLMVFGVARGVSSGKAAVELALFPYVWITLFTITII 271  
Db 205 QGI-GSPEIRWNLCLCLLAMVIVFLCLIKGVSSGKVYVFTATFPLILMLLVRCVT 263  
QY 272 LPGAIDGLFPVTPQMAKLELGVYSANVQVFSLVYCTGPIIMSSSYNGFRHNIYRDA 331  
Db 264 LPGAAMKGIQFPLTPQFHLHLSKSWIEALQIFKSLGVGFGGLLTFASYNFHONIVRDT 323  
QY 332 WIVTTLDTFSLGCTIFGLILNLAELNSEVDVAGAGTSLAFISYPAIAKTPOQ 391  
Db 324 FIVLGNATISILGAFILFSLVGLYSQELGVPV-DQAKAGPGIAFVYTPQAMMLPLSP 382  
QY 392 LFSVLEFLMSVLIGSSVALLSTFNTLMDAPF---RVPTVYSAMTSCGFLGLGYVC 448  
Db 383 FMSLFFFMMLTLGLDSQFAFLETIVAVDEFPYLRPKRAVSGGLICVAMYLMGLILT 442  
QY 449 TPGGQYILEVDHNGGFVLVFCALSELAGVFWIYGLNCLDIEFMLGKKTGAYWRLCW 508  
Db 443 TDGGMVYLVLLDDYSASFGMLVAVVITCLAVTRYGIQRCRDHMLGFKPGLYFRACW 502  
QY 509 GVTPAINTVYEFALLASNNLVFGDNVYVPTAGVSGYMLFLGMTFVPYIGIGFSLYKY 568  
Db 503 LFSPLALLLVISYKYPSEYG-SYRPPAMAILGLMGLSLCMTIPAGMLVAVLR- 560  
QY 569 RTGTFSETIKKAFHSPKSPRSPREREM 599  
Db 561 EEGSLMERLQOASRPADWGP-SLEENRTGM 590

## RESULT 4

66A7\_HUMAN STANDARD; PRT; 636 AA.  
AC 099884;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sodium-dependent proline transporter.  
GN SLC6A7 OR PROT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Hippocampus;  
RX MEDLINE=95379780; PubMed=7651355;  
RA Shafiqat S., Velaz-Faircloth M., Henzl V.A., Whitney K.D.,  
RA Yang-Feng T.L., Seidlin M.F., Fremean R.T., Jr.,  
RT "Human brain-specific L-proline transporter: molecular cloning,  
RT functional expression, and chromosomal localization of the gene in  
RT human and mouse genomes."  
RT Mol. Pharmacol. 48:219-229(1995).  
CC -1- FUNCTION: TERMINATES THE ACTION OF PROLINE BY ITS HIGH AFFINITY  
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER  
CC FAMILY (SNF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: S80071; AAB47007.2; -  
DR Genbank: HGNC:11054; SLC6A7.  
DR MIM: 606205; -  
DR InterPro: IPR000175; Na/nttran\_symport.

DR Pfam: PF00209; SNF; 1.  
DR ProDom: PD000448; Na/nttran\_symport; 1.  
DR PROSITE: PS00610; NA\_NEUTROTAN\_SYM\_1; 1.  
DR PROSITE: PS00754; NA\_NEUTROTAN\_SYM\_2; 1.  
DR PROSITE: PS50267; NA\_NEUTROTAN\_SYM\_3; 1.  
KW Neurotransmitter transporter; transport; transmembrane; glycoprotein;  
KW Symport; Amino-acid transport; Polymorphism.  
FT DOMAIN 1 45  
FT TRANSMEM 46 66  
FT TRANSMEM 74 93  
FT TRANSMEM 117 137  
FT DOMAIN 138 214  
FT TRANSMEM 215 233  
FT TRANSMEM 242 259  
FT TRANSMEM 295 312  
FT TRANSMEM 324 345  
FT TRANSMEM 378 397  
FT TRANSMEM 425 443  
FT TRANSMEM 459 479  
FT TRANSMEM 500 519  
FT TRANSMEM 538 556  
FT DOMAIN 557 636  
FT CARBOHYD 182 182  
FT VARIANT 345 345  
FT  
FT  
SQ SEQUENCE 636 AA; 70892 MM; 9A2D1071F59E409B CRC64;  
Query Match 30.4%; Score 1020; DB 1; Length 636;  
Best Local Similarity 39.8%; Pred. No. 5.4e-60;  
Matches 227; Conservative 87; Mismatches 239; Indels 18; Gaps 10;

QY 41 DTDLEAE-PPRWVSNNIEFLMSCIATSVGLGNVWRPFPAYONGGAFVYVYLL 99  
Db 26 DVDLDVDFAAHRGWMGTGIDFLSLCIGCVGLGVWHPRYAYINGGAFVYVYVLMAL 85  
QY 100 VGRVYVILECVLGGPSSRSNYSKWSISPAKGTGYQAAGCYLLSYVIGCLYLA 159  
Db 86 CGPLFFLESLGQFSSIGPLAVKISPLFGAAMLLVGLVAIYNNMIAVLYELF 145  
QY 160 MSFOATLPMAICQPEM-----ENCYPSDPTLASVNNITGTSAAOLYFRVL--QOS 211  
Db 146 ASLTSNLPWEHCGMMWNTERCLERHSPKNGALPL-NLCTVSPSEBYRSRYVLIHIGS 204  
QY 212 DGIIGGLGAPIMYLVLCFLIAMLMVFGVARGVSSGKAAVELALFPYVWITLFTITII 271  
Db 205 QGI-GSPEIRWNLCLCLLAMVIVFLCLIKGVSSGKVYVFTATFPLILMLLVRCVT 263  
QY 272 LPGAIDGLFPVTPQMAKLELGVYSANVQVFSLVYCTGPIIMSSSYNGFRHNIYRDA 331  
Db 264 LPGAAMKGIQFPLTPQFHLHLSKSWIEALQIFKSLGVGFGGLLTFASYNFHONIVRDT 323  
QY 332 WIVTTLDTFSLGCTIFGLILNLAELNSEVDVAGAGTSLAFISYPAIAKTPOQ 391  
Db 324 FIVLGNATISILGAFILFSLVGLYSQELGVPV-DQAKAGPGIAFVYTPQAMMLPLSP 382  
QY 392 LFSVLEFLMSVLIGSSVALLSTFNTLMDAPF---RVPTVYSAMTSCGFLGLGYVC 448  
Db 383 FMSLFFFMMLTLGLDSQFAFLETIVAVDEFPYLRPKRAVSGGLICVAMYLMGLILT 442  
QY 449 TPGGQYILEVDHNGGFVLVFCALSELAGVFWIYGLNCLDIEFMLGKKTGAYWRLCW 508  
Db 443 TDGGMVYLVLLDDYSASFGMLVAVVITCLAVTRYGIQRCRDHMLGFKPGLYFRACW 502  
QY 509 GVTPAINTVYEFALLASNNLVFGDNVYVPTAGVSGYMLFLGMTFVPYIGIGFSLYKY 568  
Db 503 LFSPLALLLVISYKYPSEYG-SYRPPAMAILGLMGLSLCMTIPAGMLVAVLR- 560  
QY 569 RTGTFSETIKKAFHSPKSPRSPREREM 599  
Db 561 EEGSLMERLQOASRPADWGP-SLEENRTGM 590

## RESULT 5

Db	10	DGQI-----STREVSAPVANDKPKTLVYKQKKAADLPDRDTWKGREFDL	54
Qy	62	MSGIATISVGLGVNWRPPFLAIONGGAGLIPYIVLVLLGVKPPYVYIECVIGQSSRSRSVK	121
Db	55	MSCVGAIGIIGLVNWRPPFLCGKNGGGAGLIPFLITLIPAGVPLILBCSLGQYTSIGIGL	114
Qy	122	WMSISPMKGTGAQAAGAGCYILTSYVVVICGLCYILYLMSPQATLPMALICOPW--ENCY	179
Db	115	VKMLAEMFGVGLAAAVLSFMLNITYIYIISMAITYLNSFTTLLPKQCDNPNRNDRCF	174
Qy	180	PSDPTLAASNNNTNGTSSAQLYFLRVLQSSDIGSLGAPIMYLVLCEFLIAMLYEGV	239
Db	175	SN-----YSNVTTNMTSAVVEFERNNHQMOTDGLDKP-CQIRPPLAITLAIAMILVEFC	228
Qy	240	VAGGVKSSKRAAYFLALPEYVVMITLFTITIIIPCATDGLFVTPQMAKLELGVYSA	299
Db	229	IMKGVCMWTGVVYFSATYIPYIMLILFLFRGVTLPEAKGILFYITPNERKLSDSVWLDA	288
Qy	300	VTOVFESLIVCTGPIITMFSSYNGFPHNIVRDAMIYTLIDPTFSFLSGCTIFGLIGNLAYE	359
Db	289	ATQIFPSYGLGSLIALGYSNPHNNYRDSIIYCCINCSOTSMAGFVPSIYGMAHV	348
Qy	360	LNSEVGDYVAGGTSIAFISFYSDALAKTPQPLESEVLEFLMASVVLGSSVALLSTNTL	419
Db	349	TKRSIAD-VAASGPIALFLAYEAVYQLPISPLMALFEFSLMLMGIDSOQCTVEGETA	407
Qy	420	AMAFPRV-----PTYVMSAMTSCGFLGLIVYCTPGGQYIIELDVHGTFEL-VLFCATIS	474
Db	408	LVDEYPRLLRNRELFLIAV-CLISVLIGLSNITGGIYVLRKLDYISASGMSLLFLVFF	466
Qy	475	ELAGVEMITGLEMLCIDIEFPMIGKTKGAVWRLCGVTPRAIMTVFFPALLASNNLVGD	534
Db	467	ECVSISEFGEVNFYNDIEMVGRPCIMWKLCMSEFFPIIIVAGVFIFSAAQMPPLMG-	535
Qy	535	NYVYPAGVYSGYIMFLFGMEVPIGIGISLKYRT--GTFSETIKKAFHS-----KPSW	587
Db	526	NYVFPKKGQGVGMILMSSMLIP---GIMAYMFLALGSLKQRIQWVQSEDTYAPEN	582
Qy	588	GP 589	
Db	583	GP 584	
RESULT 6			
6S61_RAT			
ID	6S61_RAT	STANDARD:	PRT: 599 AA.
AC	P23978;		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Sodium- and chloride-dependent GABA transporter 1.		
GN	SLC6A1 OR GABT1 OR GAT-1.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	TISSUE=Brain;		
RX	MEDLINE=90378307; PubMed=1975955;		
RA	Guastella J., Nelson N., Nelson H., Czyzyk L., Keynan S.,		
RA	Miedel M.C., Davidson N., Lester H.A., Kennet B.I.;		
RT	"Cloning and expression of a rat brain GABA transporter.";		
RL	Science 249:1305-1306(1990).		
RN	[2]		
RP	TOPOLOGY.		
RX	MEDLINE=97313439; PubMed=959433;		
RA	Clark J.A.;		
RT	"Analysis of the transmembrane topology and membrane assembly of the		
RT	GAT-1 gamma-aminobutyric acid transporter.";		
RL	J. Biol. Chem. 272:14695-14704(1997).		
CC	-1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY		
CC	SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.		

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR
CC STIMULANTS SUCH AS AMPHETAMINES OR COCAINE.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M59742; AAA63487.1; -.
DR PIR: A35918; ACRTGT.
DR InterPro: IPR000175; Na/ntran_symport.
DR Pfam: PF00209; SNF. 1.
DR PRINTS: PR00176; NAMEUSMPORT.
DR PRODOM: PD000448; Na/ntran_symport; 1.
DR PROSITE: PS00610; NA_NEUROTRAN_SYM_1; 1.
DR PROSITE: PS00754; NA_NEUROTRAN_SYM_2; 1.
DR PROSITE: PS50267; NA_NEUROTRAN_SYM_3; 1.
DR Neurotransmitter transport: Transport; Transmembrane; Glycoprotein;
KW Symport; Multigene family.
FT DOMAIN 1 52
FT TRANSMEM 53 73
FT DOMAIN 74 80
FT TRANSMEM 81 100
FT DOMAIN 101 123
FT TRANSMEM 124 144
FT DOMAIN 145 211
FT TRANSMEM 212 230
FT DOMAIN 231 256
FT TRANSMEM 257 291
FT DOMAIN 292 309
FT TRANSMEM 310 320
FT DOMAIN 321 342
FT TRANSMEM 343 374
FT DOMAIN 375 394
FT TRANSMEM 395 421
FT DOMAIN 422 440
FT TRANSMEM 441 456
FT DOMAIN 457 477
FT TRANSMEM 478 497
FT DOMAIN 498 517
FT TRANSMEM 518 535
FT DOMAIN 536 554
FT TRANSMEM 555 599
FT DOMAIN 176 176
FT CARBOHYD 181 181
FT CARBOHYD 184 184
SQ SEQUENCE 599 AA; 67001 MW; 4FEF85092DC1D045 CRC64;
Query Match 30.18; Score 1008; DB 1; Length 599;
Best Local Similarity 34.88; Pred. No. 3.1e-59;
Matches 211; Conservative 122; Mismatches 231; Indels 42; Gaps 13;

```

```

DB 175 SN-----YSLVNTNMTSAVVEFWRNMQHTDGLDKP-GQIRWPLAITAIWVLYFC 228
QY 240 VARGVSSGRKAAFLALEPPYVMTLTFTTITLGCARDGLFEFTPOKALEGWKSA 299
DB 229 IMKGWMTGKVVYSATYPTMLITLFRGVTLEGAEGLFTYTPNKRKISDESWLDA 288
QY 300 VTQVFFSLVCTGPIINSSYNGFRPHNIYRDAMIYTLTDPFSLGCTIGILGNLAYE 359
DB 289 ATQLEFFSGGLGLSLNAGSNHNMYNRISIVCCINSGTSMFAGFVFSYGFMAHV 348
QY 360 LNSEVDYVAGGGSIAFISIPDAIATPQPLSVLPFLMMSVLGSGSVALLSTNTL 419
DB 349 TKRSIAD-VASGGLFLAYPEAVTOLPISPLAIIFFSMLMLGIDSOQCVEGFTA 407
QY 420 AMDAFPRV-----PVTYVSAMTSCSGFLIGVYCPGGOYILELDVHGFFL-VLPCAIS 474
DB 408 LVDEYPRILNRRLRELFIAAV-CIYSLYLGSLNTIQQGYVFKLDYVSASMSLFLVVF 466
QY 475 ELAGVFWIYGLLENCLDIEFMIGKRTGAYWRICGVITPAIMTVFFVALLASNNLVFGD 534
DB 467 ECVISISMFYGNRRFYDNIQEWGSRPCIMWKLCSEFTPIIVAGVFLPSANQMTPLMG- 525
QY 535 NYVPTVIGYSGVIMLFLGMTFVPIGIFSLYKRT--GTFSETIKAFHS-----KPSW 587
DB 526 SYVFPKMGQGVGMALMSWVLP---GYNAVWFLTKSLKRLQVMIQPSEDIYVPPEN 582
QY 588 GPRSPR 593
DB 583 GPEOPQ 588

RESULT 7
ID S6A4_MACMU STANDARD; PRT: 630 AA.
AC Q9MYXO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent serotonin transporter (5HT transporter) (5HTT).
GN SLC6A4.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_Taxid=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21126513; PubMed=11223167;
RA Miller G.M., Yalin S.M., De la Garza R. II, Goulet M., Madras B.K.;
RT "Cloning of dopamine, norepinephrine and serotonin transporters from
RT monkey brain: relevance to cocaine sensitivity."
RT Brain Res. Mol. Brain Res. 87:124-143(2001).
CC -1- FUNCTION: TERMINATES THE ACTION OF SEROTONINE BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR
CC STIMULANTS SUCH AS AMPHETAMINES OR COCAINE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF285761; AA97247.1; -.
DR InterPro: IPR002437; 5HT_transporter.
DR InterPro: IPR000175; Na/ntran_symport.

```

DR Pfam; PF00209; SNF; 1.  
 DR Pfam; PF03491; SHT transporter; 1.  
 DR PRINTS; PR00176; NANUSMPORT.  
 DR ProDom; PD000448; Na/ntran\_symport; 1.  
 DR ProSITE; PS00754; NA\_NEUROTAN\_SYM\_1; 1.  
 DR ProSITE; PS00754; NA\_NEUROTAN\_SYM\_2; 1.  
 DR ProSITE; PS00754; NA\_NEUROTAN\_SYM\_3; 1.  
 DR Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
 KW Symport.  
 FT DOMAIN 1 87 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 88 108 1 (POTENTIAL).  
 FT TRANSMEM 116 135 2 (POTENTIAL).  
 FT TRANSMEM 160 180 3 (POTENTIAL).  
 FT DOMAIN 181 252 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 253 271 4 (POTENTIAL).  
 FT TRANSMEM 280 297 5 (POTENTIAL).  
 FT TRANSMEM 333 350 6 (POTENTIAL).  
 FT TRANSMEM 362 383 7 (POTENTIAL).  
 FT TRANSMEM 417 436 8 (POTENTIAL).  
 FT TRANSMEM 464 482 9 (POTENTIAL).  
 FT TRANSMEM 539 558 10 (POTENTIAL).  
 FT TRANSMEM 577 595 11 (POTENTIAL).  
 FT TRANSMEM 596 630 12 (POTENTIAL).  
 FT DOMAIN 630 70257 MW; 79DFB7E5E9EB6924 CRC64;  
 FT CARBOHYD 208 217 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 630 AA; 70257 MW; 79DFB7E5E9EB6924 CRC64;

Query Match 29.8%; Score 998; DB 1; Length 630;  
 Best Local Similarity 37.0%; Pred. No. 1.5e-58;  
 Matches 226; Conservative 98; Mismatches 241; Indels 46; Gaps 15;

1 MNDGVNGVFESSESEKPEKRRSSQSLPPANNKALNDIDPLEAE--PPRAWVSNNT 58  
 38 VESGDISNGY-SAVSPGAGDDTRHSIRAF-----TTVAELHOGREHWGKKV 86  
 59 EFLMSCATSVGLGNWRRPPIAYONGGAGFLVPYVILLVGRPVYLLGVGGSSRN 118  
 87 DELLSVIGYAVDLGNWRRPPIAYONGGAGFLVPYVILLVGRPVYLLGVGGSSRN 146  
 119 SVKWN-SISPMKNGYVQAAGCCYLLSYVYVIGLCILYLLAMSPQALIPAAICQPEW 176  
 147 CISTWRKICPTFKGIVACIATIAFYASTYNTIMAMALYLLISFTDQLPWTSCKNSMT 206  
 177 -NCVPSDPTLAASVNNIT---NGTSSAQLYFLRYLQ--QSDGIGGLGAPIMYVLCLE 230  
 207 GNC-----TNFSEENITWTLHSTPAPAEFYRHHVQLHRSKGLQ-DLGISWQALACIM 260  
 231 IATLMVEGVAVAGVSKKAAVEALFPYVNMILFTITILLPGATDGLFFVTPQMAKL 290  
 261 LIFVIVYSIMKGVTSKVVWVATFPYIILSVLLVNGATLPGAMRGVLYFLKPNMOKL 320  
 291 LELGVYSANVQVFFSLVVCGRPIIMFSSYNGFRINIRDMIVITLDTFTFLSGCTFE 350  
 321 LETGVWIDAAQIFESLSPGFCVLLAFASYNKFNCCYDALVTSVNMCTSEVSGFVLE 380  
 351 GILGNLAELNSEVDVAGAGSTSLAFISYPAIKTQPOLQSVLEFLFMSVIGSSV 410  
 381 TVLGTMALMRNDVSEVAKDAGPSLFTFYAIAINMASTFAIIFLMLTTLGLDSIF 440  
 411 ALLSTFNTLMDAFPRVPT-----VYMSAMTSCGFFLLGLVYCPGGGYILLEVHYDG 464  
 441 AGLEGVITAVLIDEPHIAKREMFVLAIVITCFGSLVTLF---GGAYVVKLLEEVAT 497  
 465 TFLVFCALSELAGFVYIGLENCLEDFEMIGKRTGAVNRCKGVITPAIMTYVFFYL 524  
 498 GPVAVLVALEAVASWFRGILQFCRDVKEMLGSPGWKRCWAISPLFLFLITCSPL 557  
 525 IASNNL-VFGDNVYPTAGYVSGYLMFLGTMFVIGIGFESLYKRTGFSETIKKAHS 583  
 558 MSPQLRLF--QYNPMSIILGYCIGISSFVCIFTYIAYKRLIS-TPGIFKRRIK----- 610  
 584 KPSWGRSPRE 594

DB 611 --STPEPTE 619.  
 RESULT 8  
 ID 561\_MUSCO STANDARD; PRT; 598 AA.  
 AC P48057;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sodium- and chloride-dependent GABA transporter 1.  
 GN SLC6A1 OR GABRI OR GAT-1.  
 OS Mus cookii (Cook's mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10098;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=95049273; PubMed=7960586;  
 RA Ruiz M., Bgal H., Sathly V.P., Qian X.J., Sarker H.K.;  
 RT "Cloning, expression, and localization of a mouse retinal gamma-  
 amino butyric acid transporter.";  
 RL Invest. Ophthalmol. Vis. Sci. 35:4039-4048(1994).  
 CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY  
 CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR  
 CC STIMULANTS SUCH AS AMPHETAMINES OR COCAINE.  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTANSMITTER SYMPORTER  
 CC FAMILY (SNF).  
 CC -----  
 CC This SNF-SYMP entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; U32178; AAA37663.1; -  
 CC MGD; MGI:95627; Gabt1.  
 DR InterPro; IPR001175; Na/ntran\_symport.  
 DR Pfam; PF00209; SNF; 1.  
 DR PRINTS; PR00176; NANUSMPORT.  
 DR ProDom; PD000448; Na/ntran\_symport; 1.  
 DR ProSITE; PS00610; NA\_NEUROTAN\_SYM\_1; 1.  
 DR ProSITE; PS00754; NA\_NEUROTAN\_SYM\_2; 1.  
 DR ProSITE; PS00754; NA\_NEUROTAN\_SYM\_3; 1.  
 KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
 KW Symport; Multigene family.  
 FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 53 73 1 (POTENTIAL).  
 FT TRANSMEM 81 100 2 (POTENTIAL).  
 FT TRANSMEM 124 144 3 (POTENTIAL).  
 FT DOMAIN 145 211 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 212 229 4 (POTENTIAL).  
 FT TRANSMEM 238 255 5 (POTENTIAL).  
 FT TRANSMEM 291 308 6 (POTENTIAL).  
 FT TRANSMEM 320 341 7 (POTENTIAL).  
 FT TRANSMEM 374 393 8 (POTENTIAL).  
 FT TRANSMEM 421 439 9 (POTENTIAL).  
 FT TRANSMEM 456 476 10 (POTENTIAL).  
 FT TRANSMEM 497 516 11 (POTENTIAL).  
 FT TRANSMEM 535 553 12 (POTENTIAL).  
 FT TRANSMEM 554 598 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 598  
 FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 598 AA; 66783 MW; 78FEC99402EC71C3 CRC64;



Query Match 29.6%, Score 992.5; DB 1; Length 598;  
 Best Local Similarity 34.5%; Pred. No. 3.3e-58;  
 Matches 209; Conservative 123; Mismatches 231; Indels 43; Gaps 13;

3 DGVNGGESESEPEMERSSQIS-LPPANKALNDIDDLAEPPERAVNSNIEFL 61  
 10 DGVN-----STVESAPASDPKPTLVKKQKAGDLPRDTWGRDPL 54

62 MSCATSVGLGNVWRPFPIAYQNGGAFVLYVLLVLPVYVLECVGFSSRSNVK 121  
 55 MSCGVAIGLGNVWRPFPIAYQNGGAFVLYVLLVLPVYVLECVGFSSRSNVK 114

122 VMSISPMKGTGVAAGCGIILSYVYVGLCLYLAMSFOATLPAICOPEN-ENCV 179  
 115 VMNAPMFKGVGLAAVLEFNLITVYIISMAIYVLSNPTTLPMKQCONPNTDRCF 174

180 PSDDTLAASVNTINGSSAOLYPLRVLOQSDIEGLAPVYVLECVGFSSRSNVK 239  
 175 SN-----YSLVNTNMTSAVEFERNHOMTDGLD--KPGQIRCLAITLAIAVLYFC 227

240 VARGVSSGKAAFLALFPYVMTLFTTTLIPATDGLFVTPOMAKLLEGVYSA 299  
 228 IMKGVMTGKAVYSATYPIMLILFRGVTLPGLAKEGLIFTTPNFKLSDSEVMDA 287

300 VTQVFEFLVCTGPIIMFSSYNGFRHNIYRDAMVITLDTFTFLSGCTIFGLGNL 359  
 288 ATQIFESYGLGLSLALGSYNSFNHNVYRDSIIVCCINCSGTFEAFVIFGMAHV 347

360 LNSEGVGAGGAGSLAFISYPAIAKTPQQLFSEVLEFMSVIGSSVALLSTNTL 419  
 348 TKRSIAD-VAAAGGLFLAPVAVTQPLSPMLALFSLMLGLIDISQPCVTEGTTA 406

420 AMDAFPRV---PTVNSAMTSCGFLGLVCTPGQYILVDHGGFTL-VLFCALS 474  
 407 LVDEYSRLNRRLFLTAAY-CIVSYLIGLSNTQGGIYVFKLFEDYSASMSLLFLVFF 465

475 ELAGVFYITGLNCLIDIEFLMKKTGAYMRLCGVTPAIAVTFVYVYALIASNNLV 534  
 466 ECVASISWFYGNRFYDIEQVSGRSPICWKSFTPIIVAGVLEFSSAVQMTPLTMG- 524

535 NYVPTAGVYGYMLFLGTFVPIGIFGSLYKRT--GTSEETIKKAFHS-----KPSW 587  
 525 SYVPRKGGVGMALMSVNLV--GYNAVYMLTLKGLKORLVMIQPSSEDIYRPN 581

588 GPRSPR 593  
 582 GPEOPQ 587

RESULT 9  
 S6A4\_HUMAN STANDARD; PRT; 630 AA.  
 ID S6A4\_HUMAN  
 AC p31645;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sodium-dependent serotonin transporter (5HT transporter) (5HTT).  
 GN SLC6A4 OR 5HTT OR SERT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC TISSUE=Brain;  
 RC TISSUE=Brain;  
 RX MEDLINE=9319744; PubMed=8452685;  
 RA Lesch K.P., Wolozin B.L., Estler H.C., Murphy D.L., Riederer P.;  
 RT "Isolation of a cDNA encoding the human brain serotonin transporter.";  
 RL J. Neural Transm. 91:67-73(1993).  
 RN [2]  
 RP MEDLINE=93211998; PubMed=7681602;  
 RA Ramamoorthy S., Bauman A.L., Moore K.R., Han H., Yang-Feng T.,

RA Chang A.S., Ganapathy V., Blakely R.D.;  
 RT "Antidepressant- and cocaine-sensitive human serotonin transporter:  
 RT Molecular cloning, expression, and chromosomal localization.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2542-2546(1993).  
 RN [3]  
 RP MEDLINE=93260476; PubMed=7684072;  
 RX MEDLINE=93260476; PubMed=7684072;  
 RA Lesch K.P., Wolozin B.L., Murphy D.L., Riederer P.;  
 RT "Primary structure of the human platelet serotonin uptake site:  
 RL Identical with the brain serotonin transporter.";  
 RN J. Neurochem. 60:2319-2322(1993).  
 RN [4]  
 RP MEDLINE=93211998; PubMed=7681602;  
 RX MEDLINE=93211998; PubMed=7681602;  
 RA Bradley C.C., Blakely R.D.;  
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: TERMINATES THE ACTION OF SEROTONINE BY ITS HIGH  
 CC AFFINITY SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR  
 CC STIMULANTS SUCH AS AMPHETAMINES OR COCAINE.  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER  
 CC FAMILY (SNP).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X70697; CA550029.1; -;  
 DR EMBL: L05568; AAA35492.1; -;  
 DR EMBL: U79746; AAB93475.1; -;  
 DR PIR: S37688; S37688.  
 DR PIR: A47398; A47398.  
 DR Genew: HGNC:11050; SLC6A4.  
 DR MIM: 182138; -;  
 DR InterPro: IPR002437; 5HT-transporter.  
 DR InterPro: IPR000175; Na/nttran-symport.  
 DR Pfam: PF00209; SNF; 1.  
 DR Pfam: PF00209; SNF; 1.  
 DR PRINTS: PR00176; NANEUSMPORT.  
 DR Prodom: PD000448; Na/nttran-symport; 1.  
 DR PROSITE: PS00610; NA\_NEUROTRAN\_SYM\_1; 1.  
 DR PROSITE: PS00734; NA\_NEUROTRAN\_SYM\_2; 1.  
 DR PROSITE: PS50267; NA\_NEUROTRAN\_SYM\_3; 1.  
 KW Neurotransmitter transporter; Transport; Transmembrane; Glycoprotein;  
 KW Symport.  
 FT DOMAIN 1 87  
 FT TRANSMEM 88 108  
 FT TRANSMEM 116 135  
 FT TRANSMEM 160 180  
 FT DOMAIN 181 252  
 FT TRANSMEM 253 271  
 FT TRANSMEM 280 297  
 FT TRANSMEM 333 350  
 FT TRANSMEM 362 383  
 FT TRANSMEM 417 436  
 FT TRANSMEM 464 482  
 FT TRANSMEM 498 518  
 FT TRANSMEM 539 558  
 FT TRANSMEM 577 595  
 FT DOMAIN 596 630  
 FT TRANSMEM 630 630  
 FT CARBOHYD 208 208  
 FT CARBOHYD 217 217  
 SQ SEQUENCE 630 AA; 70324 MW; 0EB535B0A579BDA2 CRC64;

Query Match 29.5%; Score 991; DB 1; Length 630;  
 Best Local Similarity 37.0%; Pred. No. 4.3e-58;  
 Matches 226; Conservative 97; Mismatches 242; Indels 46; Gaps 15;







```

Db 151 WRKICPFGKIGVACIACIAYIAIYNTIIMALYIISLDRPLPTSTGNSMNGNC- 209
Oy 180 PSDPTLAASVNNIT---NGTSSAOYLFLRTVLO--QSDGIEGGIGAPWLVLCFLIAML 234
Db 210 ----TNFADQNTITWHTSHTSRAEFLRHVLOHOSKGIQ-DLGTISMOLITCIVLIFT 264
Oy 235 WMEGVAVRGVKSCKAAFLALFPPVVMITLFTITLPCATDGLFFVTPOAKLLELG 234
Db 265 VYFSIMKGVKTSKGVVWMTATFPYIVLSVLVNGALPGAMKRVVFLKPNQKLETG 324
Oy 295 VWYSAVNOVFESLVCVCGIIPMFSSYNGEFRNHIYRDAMIYTTDTPFSFGCITGILG 354
Db 325 VWVDAADAIFFSLPBGCVLAFASIKFNENNCQDLVYSVNCMTSPFSGPIFYVLG 384
Oy 355 NLAYELNSEVDVYAGAGTSLAFISYDPAIAKTFQPLQFVSLFELMSVGLIGSSVALLS 414
Db 385 YMAEMRNEDSEVAKDAGPSLFTTVAEAIANMPASFFPAIFLMLITGLDSTFAGLE 444
Oy 415 TFFNLAMDAFPRVPT-----VYMSAMTSCGFLGLVYCPGGQITLLEVDHGGTFV 468
Db 445 GVITAVLDEPFIHAKRREMFVLIVITCVLGSLLT--TSGAVVVTLLSEYATGPV 501
Oy 469 LECAISELAGVFMIYGLENLCLDIEFMIGKRTGAYWRLCMGVITPAIMTVFFYALLASN 528
Db 502 LTVALLIEAVNSWFTGTFQFCSDYKEMLGFSPGMFWRLCWALISPLFLITCISFLMSP 561
Oy 529 NL-VFGDNVYVPTAGVYSGYMLFLGMEFPIGIGFSLYKRTGTFSETTICKAHKSPW 587
Db 562 QLRLP--QYNPVHMSIVYGCIGMSVYCIPTIYIYRLIS--TPGLKERIKI-----SI 612
Oy 588 GPRSPRE 594
Db 613 TPETPTE 619

```

```

RESULT 13
S6AC_RAT STANDARD: PR: 614 AA.
AC P48056:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Sodium- and chloride-dependent betaine transporter (Na+/Cl-
DE betaine/GABA transporter).
GN SLC6A12.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN 11)
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=97019277; PubMed=8665807;
RA Burnham C.E., Buerk B., Schmidt C., Bucuvalas J.C.;
RT "A liver-specific isoform of the betaine/GABA transporter in the rat:
RT cDNA sequence and organ distribution.";
RT Blochim. Biophys. Acta 1284:4-8(1996).
CC - FUNCTION: TRANSPORTS BETAINES AND GABA. MAY HAVE A ROLE IN
CC REGULATION OF GABAERGIC TRANSMISSION IN THE BRAIN THROUGH THE
CC REUPTAKE OF GABA INTO PRESYNAPTIC TERMINALS, AS WELL AS IN OSMOTIC
CC REGULATION (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - SIMILARITY: BELONGS TO THE SODIUM-NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL: U028927; AAC52867.1; ALT_INIT.
DR InterPro: IPR000175; Na+/ntran_symport.
DR Pfam: PF00209; SNF_1.
DR PRINTS: PR00176; NANEUSMPORT.
DR ProDom: PD000448; Na+/ntran_symport; 1.
DR PROSITE: PS00610; NA_NEUTROTAN_SYM_1; 1.
DR PROSITE: PS00754; NA_NEUTROTAN_SYM_2; 1.
DR PROSITE: PS0267; NA_NEUTROTAN_SYM_3; 1.
DR Neurotransmitter transporter; Transmembrane; Glycoprotein;
KW Symport.
FT DOMAIN 1 44
FT TRANSMEM 45 65
FT TRANSMEM 73 92
FT TRANSMEM 117 137
FT TRANSMEM 138 210
FT TRANSMEM 211 229
FT TRANSMEM 238 255
FT TRANSMEM 291 308
FT TRANSMEM 320 341
FT TRANSMEM 374 393
FT TRANSMEM 423 441
FT TRANSMEM 458 478
FT TRANSMEM 499 518
FT TRANSMEM 538 556
FT DOMAIN 557 614
FT CARBOHYD 171 171
FT CARBOHYD 183 183
SQ SEQUENCE 614 AA; 69748 MW; 14C78DE5E1D808B CRC64;

Query Match 29.28; Score 979; DB 1; Length 614;
Best Local Similarity 36.18; Pred. No. 2.ee-57;
Matches 216; Conservative 111; Mismatches 233; Indels 38; Gaps 13;

Oy 39 IDDTDLAEPPERPERVNNIEFLMSCIATSYGNGVWRFPIYONGGAFVPIYVLL 98
Db 25 MDQKDKD-QYKDRQMTNKRMEFVLSVAGELITGLANWRFYLYCKNGGAFPIFYFFF 83
Oy 99 LVGRPVYLLCVCVGLGQESSRNSVKW--SISPMKGTGYAQAAGCYILSYVVICGLCY 157
Db 84 SCGIPIVFLFVALGQYSQSGSVTAMRKICPLDGLGASVVISYLLIYIILANALFV 143
Oy 158 LAMSFQATLFWAICQPMW--ENCPV---SPTLAASVNNITNGTSSQQLFLRTVLOQSD 212
Db 144 LFSFTEWELPWITCTNSMNTCHCVDFLNSYSTRASYS--NFTSPVMEWERVRVIGITS 201
Oy 213 GIEGGLAPVYLYLCFIAMLVGEGVAVRGVKSCKAAFLALFPPVVMITLFTITL 272
Db 202 GIH-DLSLRKRELCLLILAMITICFIMKGVKSTGVYFTAFPIYLLILIRVTL 260
Oy 273 PGATDGLIFVTPQMAKLELGWYSAVTOVFESLVCVCGIIPMFSSYNGEFRNHIYRDAM 332
Db 261 PGAVOGIYFYLKPLDLRLKDPQVWMDAGTOIFFFAICQCLTFLGSGYHNHCYRDSI 320
Oy 333 IVTTLDFSTLSGCTIFGLIGNLAYELNSVGVGVAGAGTSLAFISYDPAIAKTFQPL 392
Db 321 ALCLFNSTSFVAQFVFSILGFMAQOGVPISE--VAESGGLFLIAFPRAVYMPISQL 379
Oy 393 FSVLEFLMSVGLIGSSVALLSTPNTLAMDAP-----RVPTVYMSAMTSCGFLIG 444
Db 380 WSCIFLMLFLPLGLDSQFVCMELVTASMDMFPQOLKRRRELLILAVAIVC--YLMG 436
Oy 445 LVYCTPGQGYLLEVDHGYGT-FLVLCFCAISELAGVMIYGLENLCLDIEFMIGKRTGAY 503
Db 437 LLVATGEGMYIFQFLDYVASSGICLLFLSEFVICWYVAGADRFYDNDVEDMIDYRPWL 496
Oy 504 WRLCMGVITPAIMTVFFYALLASNNIVEGNNVYVPTAGVYSGYMLFLGMEFPIGIGF 563
Db 497 VKISWLFLLTPDLCAATEFSSKTYPLKYNVYIYPSGISTGLWLFSSMACYPLPIII 556
Oy 564 SLVYRTGTFEETIKKAHFKSPW---GPRSPREREMQFKABAKALROKMTNTRYK 618
Db 557 TLTKIQ-GSFKRKRIQLRLTPPSLPPQGRSRPQD-----GSAQNCSTSPVK 602

```

RESULT 14  
S6A1 MOUSE STANDARD; PRT; 598 AA.

AC P31648;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sodium- and chloride-dependent GABA transporter 1.  
GN SLC6A1 OR GABT1 OR GAT-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92335351; PubMed=1631167;  
RA Liu Q.-R., Mandliyan S., Nelson H., Nelson N.;  
RL "A family of genes encoding neurotransmitter transporters.";  
Proc. Natl. Acad. Sci. U.S.A. 89:6639-6643(1992).  
CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY  
SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: BRAIN.  
CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR  
STIMULANTS SUCH AS AMPHETAMINES OR COCAINE.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM-NEUROTRANSMITTER SYMPORTER  
FAMILY (SNF).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
-----  
DR EMBL: M92377; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: M92378; -; NOT\_ANNOTATED\_CDS.  
DR PIR: F46027;  
DR MGD: MGI:95627; Gatl1.  
DR InterPro: IPR000175; Na/ntran\_symport.  
DR Pfam: PF00209; SNF; 1.  
DR PRINTS: PRO0176; NANEUSMPORT.  
DR ProDom: PD000448; Na/ntran\_symport; 1.  
DR PROSITE: PS00610; NA\_NEUROTRAN\_SYMPT\_1; 1.  
DR PROSITE: PS00754; NA\_NEUROTRAN\_SYMPT\_2; 1.  
DR PROSITE: PS0267; NA\_NEUROTRAN\_SYMPT\_3; 1.  
KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
KW Symport; Multigene family.  
FT DOMAIN 1 52  
FT TRANSMEM 53 73 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 81 100 1 (POTENTIAL).  
FT TRANSMEM 124 144 2 (POTENTIAL).  
FT DOMAIN 145 211 3 (POTENTIAL).  
FT TRANSMEM 212 229 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 238 255 4 (POTENTIAL).  
FT TRANSMEM 291 308 5 (POTENTIAL).  
FT TRANSMEM 320 341 6 (POTENTIAL).  
FT TRANSMEM 374 393 7 (POTENTIAL).  
FT TRANSMEM 421 439 8 (POTENTIAL).  
FT TRANSMEM 456 476 9 (POTENTIAL).  
FT TRANSMEM 497 516 10 (POTENTIAL).  
FT TRANSMEM 535 553 11 (POTENTIAL).  
FT DOMAIN 554 598 12 (POTENTIAL).  
FT TRANSMEM 599 617 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 176 176 N-LINKED (GLCNAC...)  
FT CARBOHYD 181 181 N-LINKED (GLCNAC...)  
FT CARBOHYD 184 184 N-LINKED (GLCNAC...)  
SQ SEQUENCE 598 AA; 66841 MW; 893C3575DB99BDD0 CRC64;

Query Match 29.2%; Score 978.5; DB 1; Length 598;  
Best Local Similarity 34.3%; Pred. No. 2.7e-57;

Matches 208; Conservative 123; Mismatches 232; Indels 43; Gaps 13;

QY 3 DQVNGGFESSEPKMKRSQIS-LPPANNKAALNDIDDTLEAEPPERAWNSNNIEPL 61  
DDB 10 DQOI-----STEVSAPVASDKPKTLTVYVOKKADLDLPRDITWKRFPDL 54  
QY 62 MSCIAITVGLGVNWRFPFLAYONGGAPLVPIVYLLGVKVVYLLGVGGFSSRNYSK 121  
DDB 55 MSCVGAIGLVNWRFPFLAYONGGAPLVPIVYLLGVKVVYLLGVGGFSSRNYSK 114  
QY 122 VMSISPMKGTGYAAGCGYILSYVYVYGLCYVILMSFPQATLPMHICPEW--ENCV 179  
DDB 115 VNNMAMPKGVAAVAALVSLFNITIIYIISAIIYILNSFTTILPMQCDNPMWTDKCF 174  
QY 180 PEDPPLAASVNNITNGTSSAQLFRTVYLOQSDGIEGIGAPIMVLYLCLFAMLVFVGV 239  
DDB 175 SN-----YSLVNTITMTSAYVEFERNHOMTDGHD--KPGIICRLATILAIYAVLYFC 227  
QY 240 VARGVSSGKAAYFLALPPYVYVNTLITITLLPGATDGIILPFYPPKAKLLEGVWSA 299  
DDB 228 IWKGVGWTKVYVESATTPYIMLITLFRGVTLPAKKGILFYITPNRKLSDSEVIFDA 287  
QY 300 VTQVEFSLVCTGPTIMSSYNGFRNHYRDAMIYTTIDTFTSFGCTIGILGNLAVE 359  
DDB 288 ATQIFESYGLGLSLIALGSYNSFNHYRDSIIYVCCINSCSMAGVIFSIVGMAYH 347  
QY 360 INSEVDVYAGGTSLATISPDIAKTFQPLESVLFLMMSVIGIGSSVALLSTFNTL 419  
DDB 348 TKRSIAD-VAAAGPLAFLAYPEAVTQDIPISPLAILFESMLMIGIDISQCTVEGFYA 406  
QY 420 AMDAPRY-----PYVMSAMTCSCGFLGLVYCPGSGGYITIELVHYGFTL-VIFCAIS 474  
DDB 407 LVDEPRLLRNRELFIANV-CIYSYLLGSLNITOGGILYTRKLPYISASGSLFLVWF 465  
QY 475 ELAGVFWYIGLEMLCLDIEFMIGKRTGAYWRLCWGVTPAINTVFFAYFALLASNNLVGCD 534  
DDB 466 ECVSISWYGVNRFEDNIOEMVGRPCIMWKLCSFFPIIVAGVFLFSAVQMPPLMG- 524  
QY 535 NVYPTAGVYSGYLMFLGKMTFVPIGIGFSLKYKT--GTSETIKKAFHS-----KPSW 587  
DDB 525 SYVFPKMGQGVWMLALSSMLIP--GYMAYMFLTKLSKORLVNIQPSIEDIVRPN 581  
QY 588 GPRSPR 593  
DDB 582 GPEQPO 587

RESULT 15  
S6AB\_HUMAN STANDARD; PRT; 632 AA.

AC P48066;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 16-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sodium- and chloride-dependent GABA transporter 3.  
GN SLC6A1 OR GABT3 OR GAT3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Fetal brain;  
RX MEDLINE=95179472; PubMed=7874447;  
RA Borden L.A., Murali Dhar T.G., Smith K.E., Branchek T.A.,  
RA Gluchowski C., Welshank R.L.;  
RT Cloning of the human homologue of the GABA transporter GAT-3 and  
RT identification of a novel inhibitor with selectivity for this site.;  
RL Recept. Channels 2:207-213(1994).  
CC -1- FUNCTION: TERMINATES THE ACTION OF GABA BY ITS HIGH AFFINITY  
SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: WIDESPREAD DISTRIBUTION IN THE BRAIN.

CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER  
CC FAMILY (SNP).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: S75989; AAB33570.1; -  
CC Gene: HGNC:11044; SLC6A11.  
CC InterPro: IPR000175; Na/nttran\_symport.  
CC Pfam: PF00209; SNF; 1.  
CC PRINTS: PR00176; NANEUSMPORT.  
CC PRODOM: PD000448; Na/nttran\_symport; 1.  
CC PROSITE: PS00610; NA\_NEUROTRAN\_SYM\_1; 1.  
CC PROSITE: PS00754; NA\_NEUROTRAN\_SYM\_2; 1.  
CC PROSITE: PS50267; NA\_NEUROTRAN\_SYM\_3; 1.  
CC Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
KW Symport; Multigene family.  
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 59 79 1 (POTENTIAL).  
FT TRANSMEM 87 106 2 (POTENTIAL).  
FT TRANSMEM 131 151 3 (POTENTIAL).  
FT DOMAIN 152 225 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 226 244 4 (POTENTIAL).  
FT TRANSMEM 253 270 5 (POTENTIAL).  
FT TRANSMEM 306 323 6 (POTENTIAL).  
FT TRANSMEM 335 356 7 (POTENTIAL).  
FT TRANSMEM 389 408 8 (POTENTIAL).  
FT TRANSMEM 438 456 9 (POTENTIAL).  
FT TRANSMEM 473 493 10 (POTENTIAL).  
FT TRANSMEM 514 533 11 (POTENTIAL).  
FT TRANSMEM 553 571 12 (POTENTIAL).  
FT DOMAIN 572 632 CYTOPLASMIC (POTENTIAL).  
FT CARBOHD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 632 AA; 70605 MW; D75C5B988B93278 CRC64;  
  
Query Match 29.2%; Score:978.5; DB 1; Length 632;  
Best Local Similarity 34.8%; Pred. No.2.9e-57;  
Matches 207; Conservative 113; Mismatches 229; Indels 45; Gaps 11;  
  
QY 22 SQQLSPANNKALNDIDDTDLAAPP-----ERRVMSNN 57  
DB 2 TAEKALPLNGKRAA-----EARESEAPGGCGSSGAAPARHPRYKRDKAVERGHWNK 56  
QY 58 IEFLLSCATSYGIGNWRPFPIAYONGGAFLLPYVIVLLVGKPYVYLCEVLGQFSSR 117  
DB 57 VEEVLSVAGEIIGLGNWRPFLLCYKNGGAFLLPYVFFICGIPVEFLETAIGQFTSE 116  
QY 118 NSVKVW-SISPMKGTGYAAGCGYILSYVVICGLCLLYLAMSFOATLPMAIQPEM- 175  
DB 117 GGITCWKRVCPLEFEGIGYATQYIEAHLNLYIIILAMALEYLSNCFTELPATCGHEMN 176  
QY 176 -ENCVPSDPTLAASVNNIT--NGTSSADLIPLRTVLOOSDGLGEGSLGAPRWLVCLFIA 232  
DB 177 TENCVEFQKLNYSNYSLSLONATSPVMEFMRVLAISDGL- HIGNLRWELACLILA 235  
QY 233 WLMVGVVARGVKSAGKAAFFLAPVYVMTFTTILPGATDGLFEFVTPQWAKLLE 292  
DB 236 WTCYFCIKGKTKGKVVYVATPTIMLLILRGVTLPGASGKIFLYLPDLSRLSD 295  
QY 293 LGVWYSAVTVQFSLVCTGPIIMSSYNGEERHNIRDAWIVTTLDTFTSLSGCTFGI 352  
DB 296 PQVWVDAGTQIFFSYAICGLTALGSYNNYNNNCYRDCIMLCINSGTSFVAGFAIFSV 355  
QY 353 LGNLALEYLSEVGDVVGAGTSLAFISPDALAKTFQPLQPSVLFELMMSVIGISVAL 412  
DB 356 LGFMAYEQGVPIAE-VAESGPGLAFLIAPKAVTKMPLSPLMWTLFFMMLIFGLGDSQFVC 414

QY 413 LSTFNFLAMDAFPRV-----PTVYMSAMTCSGFGLLGYCTPGGOYITIELVDHYGCT- 465  
DB 415 VESLVYAVDMYKRVPRRGYRRELLILALS-ISTFLGLVMTLEGGMITFQFLDSTAASG 473  
QY 466 FLVLFCAISELAGVFWIYGLLENLCLDIEFMGKRTGAYWRLCWGVTPAIMTTFEYALL 525  
DB 474 MCLLFVAIFECICIGWYGSNRFYDNIEMDIGRPPSLIKWCWMIMTPGICAGIFIEFLI 533  
QY 526 ASNNLVFGDNVYPTAGYSGYMLPLGMTFVPIGIGFSIXKRTGTFSETIKK 579  
DB 534 KYKPLKNNITYTPAMQYIGIMLALSSMLCIPLMICITVWK--TEGTLPEKLOK 586

Search completed: July 21, 2003, 09:25:20  
Job time : 26 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:24:45 ; Search time 87 Seconds  
(without alignments)  
1499.170 Million cell updates/sec

Title: US-09-991-458-2  
Perfect score: 3354  
Sequence: 1 MNDGVNGGFESSEPKMEPK.....TSRYKHLWYSGAYRRIN 633

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.protent:\*  
12: sp.virus:\*  
13: sp.verticillate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriaph:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3354	100.0	633	5 Q9U5A9	Q9U5A9 manduca sex
2	3084	91.9	634	5 Q76188	Q76188 manduca sex
3	1533	45.4	593	5 Q9W4C5	Q9W4C5 drosophila
4	1510.5	45.0	654	5 Q95T51	Q95T51 drosophila
5	1492	44.5	651	5 Q9V5B3	Q9V5B3 drosophila
6	1472.5	43.9	639	5 Q9VJR4	Q9VJR4 drosophila
7	1461.5	43.6	739	5 Q9NKA8	Q9NKA8 drosophila
8	1414	42.2	629	5 Q9V690	Q9V690 drosophila
9	1389.5	41.4	593	5 Q9VSV2	Q9VSV2 drosophila
10	1238.5	36.9	615	5 Q8T7E1	Q8T7E1 aedes aegypti
11	1089.5	32.5	791	11 Q91ZQ2	Q91ZQ2 mus musculus
12	1067.5	31.8	415	5 Q8SVT9	Q8SVT9 drosophila
13	1030	30.7	638	11 Q9JMA9	Q9JMA9 mus musculus
14	1030	30.7	674	5 Q8WPM9	Q8WPM9 oikopleura
15	1027	30.6	638	11 Q9D317	Q9D317 mus musculus
16	1025	30.6	638	11 Q91Y60	Q91Y60 mus musculus

17	1024	30.5	638	11 Q9R183	Q9R183 mus musculus
18	1010	30.1	642	4 Q9UN76	Q9UN76 homo sapien
19	1003.5	29.9	598	13 Q37771	Q37771 raja sp. 9a
20	999	29.8	598	13 Q91494	Q91494 torpedo cal
21	997.5	29.7	631	6 Q9GMA5	Q9GMA5 ovis aries
22	979	29.2	611	13 Q91502	Q91502 torpedo mar
23	971	29.0	581	5 Q23969	Q23969 drosophila
24	961	28.7	614	11 Q8VCS9	Q8VCS9 mus musculus
25	961	28.7	630	13 Q42482	Q42482 rana catesb
26	960	28.6	640	11 Q8VBM1	Q8VBM1 mus musculus
27	956	28.5	602	4 Q8TCC2	Q8TCC2 homo sapien
28	956	28.5	625	13 Q91BH6	Q91BH6 cyprinus ca
29	954.5	28.5	622	13 Q91503	Q91503 torpedo mar
30	950	28.3	629	13 Q9DE16	Q9DE16 oreochromis
31	949.5	28.3	621	11 Q91W12	Q91W12 mus musculus
32	947.5	28.2	629	13 Q90ZY1	Q90ZY1 brachydanio
33	947.5	28.2	1201	5 Q9W107	Q9W107 drosophila
34	945	28.2	633	11 Q8VC47	Q8VC47 mus musculus
35	944	28.1	635	13 Q9DGN5	Q9DGN5 gallus gall
36	942.5	28.1	620	6 Q9GJ76	Q9GJ76 macaca fasc
37	939	28.0	619	11 Q9J341	Q9J341 mus musculus
38	938.5	28.0	620	6 Q9GJ75	Q9GJ75 salmuri sci
39	938	28.0	617	11 Q63380	Q63380 rattus norv
40	936.5	27.9	620	6 Q9GJ77	Q9GJ77 salmuri sci
41	934.5	27.9	597	11 Q9WTR4	Q9WTR4 rattus norv
42	934	27.8	619	11 Q9R0X6	Q9R0X6 mus musculus
43	931.5	27.8	617	11 Q8R212	Q8R212 mus musculus
44	929.5	27.7	567	11 Q9WTR3	Q9WTR3 rattus norv
45	929.5	27.7	597	5 Q25512	Q25512 manduca sex

## ALIGNMENTS

RESULT 1  
ID Q9U5A9 PRELIMINARY: PRT; 633 AA.  
AC Q9U5A9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Amino acid transporter/amino acid-gated channel for sodium/potassium ions.  
GN CAATCH1.  
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; OC Sphingidae; Sphingidae; Sphinginae; Manduca.  
OX NCBI\_TaxID=7130;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=WIDGUT;  
RX MEDLINE=20387341; PubMed=10829035;  
RA Feldman D.H., Harvey W.R., Stevens B.R.;  
RT "A novel electrogenic amino acid transporter is activated by K+ or Na+, is alkaline pH-dependent, and is Cl--independent.";  
RL J. Biol. Chem. 275:24518-24526(2000).  
DR EMBL: AF013963; AAF18560.1; -;  
DR InterPro: IPR000175; Na/ntran\_symport.  
DR Pfam: PF00209; SNE; 1.  
DR PRINTS: PR00176; NANUSMPORT.  
DR PRODOM: PD000448; Na/ntran\_symport; 1.  
DR PROSITE: PS50267; NA\_NEUTRAN\_SYMPT\_3; 1.  
SQ SEQUENCE 633 AA; 69934 MW; 597EDC58E88FDFID CRC64;

Query Match 100.0%; Score 3354; DB 5; Length 633;  
Best local Similarity 100.0%; Pred. No. 1.4e-259;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNDGVNGGFESSEPKMEPKRSSQISLPANKKALNDIDDDTDLAEPPERWVMSNNIEF 60  
|||||  
DB 1 MNDGVNGGFESSEPKMEPKRSSQISLPANKKALNDIDDDTDLAEPPERWVMSNNIEF 60  
|||||





RESULT 5	Q9V5B3	PRELIMINARY:	PRT;	651 AA.
ID	Q9V5B3			
AC	Q9V5B3;			
DT	01-MAY-2000	(TREMBLrel_13, Created)		
DT	01-MAY-2000	(TREMBLrel_13, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel_21, Last annotation update)		
DE	CG1698 protein.			
GN	CG1698.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_taxonomy:7227;			
LN	[1]			

PC  
RP  
SEQUENCE FROM N.A.  
RX STRAIN-BERKELEY;  
RX MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeell M.D., Zheng Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blasei R.C., Chaipe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Chaipe M., Pfeiffer B.D.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale B.J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bentman B.P., Bhandari D., Bolashkov S.,  
RA Borjova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke S., Dahlen C., Davenport L.B., Davies P.,  
RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,  
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glaser A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kreft C., Kravitz S., Kulp D., Lai Z.,  
RA Liou X., Mattel B., McIntosh C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreft A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,  
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Stryker S., Tector A.C., Stapleton M., Strong R., Sun E.,  
RA Wang Z.-Y., Wasatman D.A., Weinstock G.M., Wang A.H., Wang X.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
DR EMBL: AEO03832; AAF58802.1; -  
DR FlyBase: Fgana0033443; CG1698.  
DR InterPro: IPRO00175; Na/atran\_sympart.  
DR Pfam: PF00209; SNE; 1.  
DR PRINTS; PRO0176; NANEUSMP0RT.  
DR PRODom: PD000448; Na/atran\_sympart; 1.  
DR PROSITE: PSS0267; NA\_NEUTRAN\_SYMP\_3; 1.  
SQ SEQUENCE 651 AA, 72070 MW, 30C81A13C973285 CRC64;

Query Match	44.5%;	Score 1492;	DB 5;	Length 651;
Best Local Similarity	48.4%;	Pred. No. 7.7e-111;		
Matches 305;	Conservative 96;	Mismatches 101;		

20; mismatches 191; indels 38; Gaps 12,

Db

38 ESSNGKRTLSEGGSGAPPGGHDS-----PSGYEAGDQPKKRDSNNNIEFLMSCTALS 91

69 GGLGNWRFEPFLAYONGGCAFLVPYVYIYLLNGKPPVYVECVLQGFSSRNSVKVWSISPA 128  
 92 VGLGNWRFEPFLAYONGGCAFLVPYVYIYLLNGKPPVYVECVLQGFSSRNSVKVWSISPA 128

[illegible]

188 SVNNITNGTSSAQLYFLRTVLQOSDGEIGGCGAPIMWLVLCETIAMLWVEGVVARGVKS 247  
::|  
QY

212 RATSLAGCGCCTHSFPIRNIIIRKASIDDDGIGYPSWSLALALAVAVITVAGINEKGVKSS 271

272 GKASIFLALFPYVMVLIVKALTLPGAFDGLYELRPFQWKRLLTEPQWYAAVATQVEEFL 331

```

308 IVCGPIIERSSYNGFRHNHYRDAMIVTITDFTSLSGCIIFGLIGNLAYELN-SEVD 366
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
332 AICGNIIVASYNGRGNHYRDAMIVTITDFTSLSGVIIGTICNIAYENNRIRIS 390
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

367 WGAGGTSLAFISYPDAIAK-TFOQLSVLFFLMMSVLGIGSSVALLSTFNTLADAPF 425

426 RVPVMSAMTSCGFLGLVYCYCPGGQYILEVDHYGGTFGLVLPFCALISELAGVEWYGL 485  
522 VVN GGFGLAF ISTDALAKFKWLPQLFSVLFLEMLFVLICGSNNMGASCMSTVINDQFG 450

451 HEKNMTVVGIAIVGFLGLLTYTPGGGPELLNVDYEGVTFVALVLAIFEEVLTAMITGV 510  
486 ENICIDTEEMTGGRTGAVWDYCIQVETNA

511 KRLCRDEEMIGIKTSLERYRICAWAVTPELMLTLIYTLVLVEPLKRYD-YTYGSGVYVF 569

546 GILMLFGMTFPIGIGISLY-----KYTGTESEETIKAFHKSQWSPRPRER 596  
Y : : : :  
570 GWCL-----SAFGVGOVLFNALPAVRKOPSHRGITLADTFFDFTPTHTCCCT--

**Y**

597 EMOFKAE-AKALROKMNTSRVKNLWYSI 624  
| | | : :  
: : : :

021

022 RIJLVOEGNANALFRSS-----IWHKI 645

09VJR4  
D 09VJR4  
09VTD4.  
PRELIMINARY; PRT; 639 AA.

01-MAY-2000 (TREMBlrel. 13, Created)
01-MAY-2000 (TREMBlrel. 13, Last Sequence update)
01-JUN-2002 (TREMBlrel. 21, Last Sequence update)

BC:DS03431.1 protein (Gn16161p).  
 BC:DS03431.1 OR CG5279.  
 Drosophila melanogaster (Fruit fly).

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea: Drosophilidae: Drosophila.  
OX NCBI\_TaxID=7227,  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Baas J., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeagwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Lau P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattali B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Stimpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Weiss A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodedge T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gbds R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL [2]  
RN SCIENCE FROM N.A.  
RP STRAIN-BERKELEY;  
RC Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Fafan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nyuno C., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AE003645; AAF3415.1; -  
DR EMBL; AY058396; AAL1625.1; -  
DR FlyBase; FBgn0028886; BG:DS0331.1.  
DR InterPro; IPR000175; Na/nttran\_sympart.  
DR Pfam; PF00209; SNF.1.  
DR PRINTS; P000176; NAMEUSMPRT.  
DR PRODOM; PD000448; Na/nttran\_sympart; 2.  
DR PROSITE; PS50267; NA\_NEUTOTRAN\_SYMP\_3; 1.  
QO SEQUENCE 639 AA; 71400 MT; B36CE08DF73DCE05 CRC64;

[illegible][illegible]

RESULT 7			
ID	Q9NKA8	PRELIMINARY;	PRT; 739 AA.
AC	Q9NKA8;		
DT	01-OCT-2000 (TREMBLrel, 15, Created)		
DT	01-OCT-2000 (TREMBLrel, 15, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel, 21, Last annotation update)		
DE	Hypothetical 82.8 kDa protein.		
GN	BG:DS03431.1 OR CG15279.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OX	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid=7227;		
RI	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=BERKELEY;		
RC	MEDLINE=99403001; PubMed=10471707;		
RA	Asburner M., Mista S., Roote J., Lewis S.E., Blazer R., Davis T.,		
RA	Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,		
RA	Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,		
RA	Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitlaw K.,		
RA	Seinkner S., Rubin G.M.;		
RT	"An exploration of the sequence of a 2.9-Mb region of the genome of		
RT	Drosophila melanogaster: the Adh region.";		
RL	Genetics 153:179-219(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RA	Seinkner S.E., Agbayani A., Arcaina T.T., Baxter E., Blazer R.G.,		
RA	Butenhoff C., Champagne M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,		
RA	Fafian D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,		
RA	Houston K.A., Hummasti S.R., Kaira K., Kearney L., Kim E., Lee B.,		
RA	Lewis S., Li P., Lomocan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,		







OY 572 TFSSTIKAFHSPKSPRREREMOKFAKALROK 611  
 DB 537 SLKSRINNSIKPHSDMCPSPDKLMDYOMF-----LRNK 570

## RESULT 10

ID 087E1 PRELIMINARY; PRT; 615 AA.

AC 087E1; 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Amino acid transporter.  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Aedes.  
 OX NCBI\_TaxID=7159;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dasher M.K., Kohn A.B., Harvey W.R., Stevens B.R.;  
 RT "AENART, A Novel Amino Acid Transporter from Larval Aedes aegypti  
 RT Midgut.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF369383; AAA13400.1;  
 SQ SEQUENCE 615 AA; 68985 MW; E14A5ED2E809A139 CRC64;

Query Match 36.98; Score 1238.5; DB 5; Length 615;  
 Best Local Similarity 46.08; Pred. No. 1,3e-90;  
 Matches 267; Conservative 79; Mismatches 173; Indels 61; Gaps 12;

OY 76 RPEPIATONGGATFVYVILLVGRPVYLLVGLGQSSRSNYSVSPAMKGTGYA 135  
 DB 44 RPEPIATONGGATFVYVILLVGRPVYLLVGLGQSSRSNYSVSPAMKGTGYA 103  
 OY 136 QACGCGYILSYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 191  
 DB 104 QLVTVTSVYVYVYLLVGLGQSSRSNYSVSPAMKGTGYA 161  
 OY 192 ITNG--TSSAQLVFLRVYLLVGLGQSSRSNYSVSPAMKGTGYA 249  
 DB 162 STSQKQVSSQIYLLVGLGQSSRSNYSVSPAMKGTGYA 221  
 OY 250 AAYFLALFPYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 309  
 DB 222 VAYFLALFPYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 281  
 OY 310 CTGPIIMSSYNGRHNRYRDAMVYVLLVGLGQSSRSNYSVSPAMKGTGYA 369  
 DB 282 GMSGIIIMFSSYNGRHNRYRDAMVYVLLVGLGQSSRSNYSVSPAMKGTGYA 341  
 OY 370 AGGTSIAFISYPAIAK-TEQPOLFSVFLFMSVGLIGSSVALLSTFNTLAMDAPRPV 428  
 DB 342 KSGGIAFISYPAIAKFDIYVLPQFVFLFMSVGLIGSSVALLSTFNTLAMDAPRPV 401  
 OY 429 TVYSAMTSCGCGFLGLVY--CTPGGQYILEVDHYGTFVLFCAISELAGVWYIGLE 486  
 DB 402 YMHALLISTIGFTGLVYIHLVANGYSIMSTTE--AFSDRLAITEWVAIWIYIGLD 458  
 OY 487 NLCDIEFMGLKKGAVWRLCMGVITPAIMTTFVYVALLANNLVFGD----- 534  
 DB 459 NWCNDIEFMVORVGLY-----VAAMLGSSNMTFHCGRHLELVYK 501  
 OY 535 ----NYVPTAGVYSGYLMFLGATFVIGIGFSLYKRTGTFSE-----TIKAFHSP 585  
 DB 502 PTYGGQYVPGALIGIFVIFIGLQVLYLVAMVTV---TGDSTRKESWGMISKAAPSS 557  
 OY 586 SWGPRSPRREREMOKFAKALRO-----KMTSRVXKL 620  
 DB 558 QMCPCTEHTKRAMLKYKEAKSRDDIYAKNHSSIVAKL 597

RESULT 11

O91202  
 ID 091202 PRELIMINARY; PRT; 791 AA.

AC 091202; 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Glycine transporter type 2.  
 GN GLYT2.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BALB/C; TISSUE-BRAIN;  
 RA Liu Q.-R., Li Q.-F.;  
 RT "Cloning and expression of mouse glycine transporter 2 (Glyt2)."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF411042; AAL17054.1;  
 DR InterPro: IPR001920; Asp/Glu\_race.  
 DR InterPro: IPR001920; Na/Na+trn\_symp.  
 DR Pfam: PF00209; SNF\_1.  
 DR Prodom: PD000448; Na/Na+trn\_symp.  
 DR PROSITE: PS00923; ASP\_GLU\_RACEMASE\_1; UNKNOWN\_1.  
 DR PROSITE: PS00610; NA\_NEUROTRAN\_SYMP\_1; UNKNOWN\_1.  
 DR PROSITE: PS00754; NA\_NEUROTRAN\_SYMP\_2; 1.  
 DR PROSITE: PS0267; NA\_NEUROTRAN\_SYMP\_3; 1.  
 SQ SEQUENCE 791 AA; 87029 MW; 8AABB600D9EAEF CRC64;

Query Match 32.58; Score 1089.5; DB 11; Length 791;  
 Best Local Similarity 37.68; Pred. No. 1,4e-78;  
 Matches 225; Conservative 105; Mismatches 229; Indels 39; Gaps 9;

OY 35 ALNIDDTDEAPPRERWVSNIEFLMCIATSVGIGVWVREPIATONGGATFVYV 94  
 DB 169 ATTTTPEDEGDEKARKGNSSKLDPLTMSVAGVAGLGNWRRPIYALFONGGATFVYV 228  
 OY 95 IVLLVGRPVYLLVGLGQSSRSNYSVSPAMKGTGYAQAAGCGYILSYVYVIGLCL 154  
 DB 229 MMLALGLPIFLFVLSIGQFASOGPVSWKAIPALOGCGIAMLISVLAIVYVITCYT 288  
 OY 155 LYTLAMFOATLPAICPEW-----ENCYSPD-----PRT 185  
 DB 289 LFTLFASFVSVLPMWGSNNMWNTEBECKDKTKLLDSCVIDDHKIQIKNSFCMTAVPNL 348  
 OY 186 A-ASVNNITNGT--SSAQLVFLRVYLLVGLGQSSRSNYSVSPAMKGTGYA 242  
 DB 349 TVNFTSQTKTFVSGSEEFKFKVLAISGIEP-GEIRPLAFCLFLAMVIVYASLAK 407  
 OY 243 GYSSGKAAYFLALFPYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 DB 408 GIKSSGKAYFLALFPYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 369  
 OY 303 VFSFLVCTPIIMFSSYNGRHNRYRDAMVYVLLVGLGQSSRSNYSVSPAMKGTGYA 362  
 DB 468 IFSLSAMKGLITLSSYNFHNHCYRDLIVCTNATISIFAGVYFISYIGMANERKV 527  
 OY 363 EGVGVAGGTSIAFISYPAIAKFTQPOLFSVFLFMSVGLIGSSVALLSTFNTLAMD 422  
 DB 528 NIENVADO-OPGIAFVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 OY 423 AFRPVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 DB 587 EFRPVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 OY 480 FVYVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 DB 647 SYVYVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 OY 540 TAGVYVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302  
 DB 706 NMSVYVYVYVYVYVIGLCLYLLMSPOATLPAICPEW-ENCYSPDPLTAAV---NN 302

Query Match	31.8%;	Score 1067.5;	DB 5;	Length 415;
Best Local Similarity	52.9%;	Pred. NO. 3.7e-77;		
Matches 199;	Conservative 73;	Mismatches 97;	Indels 7;	Gaps 5

Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi, Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

Query Match	30.7%;	Score 1030;	DB 11;	Length 638;
Best Local Similarity	38.5%;	Pred. No. 6e-74;		
Matches 230; Conservative	93;	Mismatches 225;	Indels 50;	Gaps 14;

[illegible]

	RESULT	14
OBWPM9	PRELIMINARY:	PRT: 674 AA.
OBWPM9		
01-MAR-2002 (TrEMBLrel. 20, Created)		
01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
Similar to glycine transporter.		
BAC001.6.		
Okkopleura diolca.		
Eukaryota; Metazoa; Chordata; urochordata; Appendicularia;		



Db 506 IEMMIGAKRMIFWLMWRACWFIPTPILLSAILWMSLVKFRHPDYAD-IPYDVGVALGWC 564  
Qy 549 MLFLGMTFVPJIGIGFSLYK--YRTGFSEETIKKAFHSHKPSWGPSPRPREREMQPKAE 604  
Db 565 MIIICIIWIFI--MAIKIYQAEENIILORITSCCIPASNMGPYLEKHERERYDMAE 619

Search completed: July 21, 2003, 09:27:47  
Job time : 91 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:24:45 ; Search time 18 Seconds  
(without alignments)  
1034.706 Million cell updates/sec

Title: US-09-991-458-2

Perfect score: 3354

Sequence: 1 MNOGVNGGVSSSEPKMEPK.....TSRYKHLMTGAYRRNIN 633

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PT05.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description
	1	1084.5	32.3	797	2	US-08-700-013B-21	Sequence 21, Appl
	2	1082.5	32.3	797	4	US-09-182-728A-2	Sequence 2, Appl
	3	1082.5	32.3	797	4	US-09-795-232-2	Sequence 2, Appl
	4	1080.5	32.2	797	2	US-08-700-013B-19	Sequence 19, Appl
	5	1080.5	32.2	797	4	US-09-191-468-122	Sequence 122, App
	6	1080.5	32.2	797	4	US-09-191-468-124	Sequence 124, App
	7	1076.5	32.1	797	4	US-09-191-468-120	Sequence 120, App
	8	1076.5	32.1	799	2	US-08-700-013B-27	Sequence 27, Appl
	9	1030.5	30.7	667	1	US-07-879-617A-8	Sequence 8, Appl
	10	1030.5	30.7	667	1	US-08-753-985-8	Sequence 8, Appl
	11	1017	30.3	635	1	US-07-879-617A-9	Sequence 9, Appl
	12	1017	30.3	635	1	US-08-753-985-9	Sequence 9, Appl
	13	1011	30.1	599	1	US-08-301-722A-5	Sequence 5, Appl
	14	1008	30.1	599	1	US-08-295-814E-11	Sequence 11, Appl
	15	1008	30.1	599	1	US-08-240-783B-4	Sequence 4, Appl
	16	1008	30.1	599	3	US-09-084-813-4	Sequence 4, Appl
	17	1008	30.1	599	4	US-09-343-361-11	Sequence 11, Appl
	18	1008	30.1	599	5	PCT-US92-09662-4	Sequence 4, Appl
	19	991	29.5	630	1	US-07-959-943-11	Sequence 11, Appl
	20	983	29.3	599	1	US-07-879-617A-10	Sequence 10, Appl
	21	983	29.3	630	1	US-08-753-985-10	Sequence 10, Appl
	22	981	29.2	630	1	US-07-959-943-9	Sequence 9, Appl
	23	978.5	29.2	632	4	US-08-295-814E-10	Sequence 10, Appl
	24	978.5	29.2	632	4	US-09-343-361-10	Sequence 10, Appl
	25	973	29.0	632	5	PCT-US93-01959-10	Sequence 10, Appl
	26	973	29.0	627	4	US-08-295-814E-4	Sequence 4, Appl
	27	973	29.0	627	4	US-09-343-361-4	Sequence 4, Appl

28	973	29.0	627	5	PCT-US93-01959-4	Sequence 4, Appl
29	971.5	28.9	607	1	US-07-959-943-7	Sequence 7, Appl
30	968.5	28.9	653	1	US-07-782-298-2	Sequence 2, Appl
31	967	28.8	687	3	US-08-834-467-2	Sequence 2, Appl
32	967	28.8	687	4	US-09-336-177-2	Sequence 2, Appl
33	965	28.8	602	1	US-08-295-814E-2	Sequence 2, Appl
34	965	28.8	602	4	US-09-343-361-2	Sequence 2, Appl
35	965	28.8	602	5	PCT-US93-01959-2	Sequence 2, Appl
36	962	28.7	627	1	US-08-291-299-10	Sequence 10, Appl
37	962	28.7	627	5	PCT-US95-10579-10	Sequence 10, Appl
38	960.5	28.6	607	1	US-07-879-617A-12	Sequence 12, Appl
39	960.5	28.6	607	1	US-08-753-985-12	Sequence 12, Appl
40	958	28.6	614	1	US-08-291-299-7	Sequence 7, Appl
41	958	28.6	614	5	PCT-US95-10579-7	Sequence 7, Appl
42	953.5	28.4	621	1	US-08-295-814E-6	Sequence 6, Appl
43	953.5	28.4	621	4	US-09-343-361-6	Sequence 6, Appl
44	953.5	28.4	621	5	PCT-US93-01959-6	Sequence 6, Appl
45	949.5	28.3	620	1	US-08-301-722A-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-700-013B-21

Sequence 21, Application US/08700013B

Patent No. 5919653

GENERAL INFORMATION:

APPLICANT: Albert, Vivian R.

APPLICANT: Kowalski, Leslie R.Z.

APPLICANT: Borden, Laurence A.

APPLICANT: McKelvy, Jeffrey F.

TITLE OF INVENTION: Human Glycine Transporter

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700.013B

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 317743-108

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3255

TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 797 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

Query Match

Best Local Similarity 32.3%: Score 1084.5; DB 2;

Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10;

QY 54 WSNNIEFLMSCIATSVGLGNVWREPFITAYONGGAFIVYVILLVGVKPYVYECVLGQ 113  
114 FSSNSVKVMSISPRAMGTGTAOANGCYIISTYVVICGLCTYTLAMSFOATLPMACOP 173  
Db 194 WSKKIDFLISMGVYAVGLGNVWREPFITAYONGGAFILIPYLMALAGLPDIFLEVLGQ 253  
QY 114 FSSNSVKVMSISPRAMGTGTAOANGCYIISTYVVICGLCTYTLAMSFOATLPMACOP 173  
Db 254 FASGPPVSWKAIPLDGGCGIAMLINSVLAIYVNIICYTLFYLFSFVSFLPWGSCNN 313  
QY 174 EM-----ENCVPSD--PTL-----AASVNIT--NGTSSA-----OL 201  
Db 314 PMNPECKDKTKLLDSCVISHDKIQKNSFCMTAPNVTMNFISQANKTVSSGSEE 373  
QY 202 YFLRTVYQOQSDGICGIGAPIWYLVLCFLTAMLVFQVAVGKSSGKAAYFLALFPYV 261  
Db 374 YFKYFVLKISAGIYYP-GEIRMPALCLFLAMVIVYASLAKGITSKVYVFTATFPYV 432  
QY 262 MITLFTIILPGATDGLFVFPQMAKLELGWYSAVOVPSLVCGPIIMESSYN 321  
Db 433 LVILLINGVILPGAGAIWFTITPKWEKLDATYWKDAATQIFSLAAMGGLTTLSSYN 492  
QY 322 GFRHNIYRDAMVITTLDTFTSFLSGCTIFGILGNLAVELNSEVDVYAGGTSIAFTSY 381  
Db 493 KFHNNCYRDLIVCTNSATSIFFAGFVIFSVIGFMANERKNIENVADQ--GPIAFVYV 551  
QY 382 DAIKTOPOLFSVLEFLMSVLCIGSSVALLSTFNLMADAPRPVTVMSANT---CS 438  
Db 552 EALTRPLSPFWALIFFLMLTLGLDTMFATIEIVTISIDEPKYLRTHKPVFTLCCV 611  
QY 439 CGFLGLVYCTPGGOYILEVDHYSGTFVLFCALISLAVGFMYIGLENCLDIEFMTGK 498  
Db 612 CFPIFGPMITOGGIVMFOLVDYTAASYALYIAIFELVIGISYVIGQRCEDENMTIGF 671  
QY 499 KTAGVWRLCNGVITPAIMTVEFYALLASNNLVFGDNVYVPTAGVSGYLMFLGMTFV 558  
Db 672 QPNIFWKCWAFTPTILFTLFCFSFYQMEPMYTG--SYRYPNMSMVLGMLACSVIWP 730  
QY 559 IGIGFLXKXRTGTFESTIKKAFHSKPSMGPRSPRERE 597  
Db 731 IMFVIKMH-LAPGRFIERLKLSPQPDWGPFLAQHGE 768

RESULT 2  
US-09-182-728A-2  
; Sequence 2, Application US/09182728A  
; Patent No. 6238883  
; GENERAL INFORMATION:  
; APPLICANT: BROWN, ANTHONY  
; APPLICANT: CHAPMAN, CONRAD GERALD  
; APPLICANT: GLOGER, ISRAEL SIMON  
; APPLICANT: EVANS, JOANNE RACHEL  
; APPLICANT: CAIRNS, WILLIAM  
; APPLICANT: HERDON, HUGH  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30176  
; CURRENT APPLICATION NUMBER: US/09/182,728A  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 981889.7  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Vets on 3.0  
; SEQ ID NO 2  
; LENGTH: 797  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-182-728A-2

Query Match 32.3%; Score 1082.5; DB 4; Length 797;  
Best Local Similarity 38.3%; Pred. No. 1.3e-93;  
Matches 222; Conservative 105; Mismatches 213; Indels 39; Gaps 10;

QY 54 WSNNIEFLMSCIATSVGLGNVWREPFITAYONGGAFIVYVILLVGVKPYVYECVLGQ 113  
114 FSSNSVKVMSISPRAMGTGTAOANGCYIISTYVVICGLCTYTLAMSFOATLPMACOP 173  
Db 194 WSKKIDFLISMGVYAVGLGNVWREPFITAYONGGAFILIPYLMALAGLPDIFLEVLGQ 253

QY 114 FSSNSVKVMSISPRAMGTGTAOANGCYIISTYVVICGLCTYTLAMSFOATLPMACOP 173  
Db 254 FASGPPVSWKAIPLDGGCGIAMLINSVLAIYVNIICYTLFYLFSFVSFLPWGSCNN 313  
QY 174 EM-----ENCVPSD--PTL-----AASVNIT--NGTSSA-----OL 201  
Db 314 PMNPECKDKTKLLDSCVISHDKIQKNSFCMTAPNVTMNFISQANKTVSSGSEE 373  
QY 202 YFLRTVYQOQSDGICGIGAPIWYLVLCFLTAMLVFQVAVGKSSGKAAYFLALFPYV 261  
Db 374 YFKYFVLKISAGIYYP-GEIRMPALCLFLAMVIVYASLAKGITSKVYVFTATFPYV 432  
QY 262 MITLFTIILPGATDGLFVFPQMAKLELGWYSAVOVPSLVCGPIIMESSYN 321  
Db 433 LVILLINGVILPGAGAIWFTITPKWEKLDATYWKDAATQIFSLAAMGGLTTLSSYN 492  
QY 322 GFRHNIYRDAMVITTLDTFTSFLSGCTIFGILGNLAVELNSEVDVYAGGTSIAFTSY 381  
Db 493 KFHNNCYRDLIVCTNSATSIFFAGFVIFSVIGFMANERKNIENVADQ--GPIAFVYV 551  
QY 382 DAIKTOPOLFSVLEFLMSVLCIGSSVALLSTFNLMADAPRPVTVMSANT---CS 438  
Db 552 EALTRPLSPFWALIFFLMLTLGLDTMFATIEIVTISIDEPKYLRTHKPVFTLCCV 611  
QY 439 CGFLGLVYCTPGGOYILEVDHYSGTFVLFCALISLAVGFMYIGLENCLDIEFMTGK 498  
Db 612 CFPIFGPMITOGGIVMFOLVDYTAASYALYIAIFELVIGISYVIGQRCEDENMTIGF 671  
QY 499 KTAGVWRLCNGVITPAIMTVEFYALLASNNLVFGDNVYVPTAGVSGYLMFLGMTFV 558  
Db 672 QPNIFWKCWAFTPTILFTLFCFSFYQMEPMYTG--SYRYPNMSMVLGMLACSVIWP 730  
QY 559 IGIGFLXKXRTGTFESTIKKAFHSKPSMGPRSPRERE 597  
Db 731 IMFVIKMH-LAPGRFIERLKLSPQPDWGPFLAQHGE 768

RESULT 3  
US-09-795-232-2  
; Sequence 2, Application US/09795232  
; Patent No. 6426405  
; GENERAL INFORMATION:  
; APPLICANT: Anthony M. Brown  
; APPLICANT: Conrad Gerald Chapman  
; APPLICANT: Israel Simon Gloger  
; APPLICANT: Joanne Rachel Evans  
; APPLICANT: William Cairns  
; APPLICANT: Hugh Jonathan Herdon  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30176-D1  
; CURRENT APPLICATION NUMBER: US/09/795,232  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 09/182,728  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 981889.7  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 797  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-795-232-2

Query Match 32.3%; Score 1082.5; DB 4; Length 797;  
Best Local Similarity 38.3%; Pred. No. 1.3e-93;  
Matches 222; Conservative 105; Mismatches 213; Indels 39; Gaps 10;

QY 54 WSNNIEFLMSCIATSVGLGNVWREPFITAYONGGAFIVYVILLVGVKPYVYECVLGQ 113  
114 FSSNSVKVMSISPRAMGTGTAOANGCYIISTYVVICGLCTYTLAMSFOATLPMACOP 173  
Db 194 WSKKIDFLISMGVYAVGLGNVWREPFITAYONGGAFILIPYLMALAGLPDIFLEVLGQ 253



```

/ TELEFAX: 609-520-3259
/ TELFX:
/ INFORMATION FOR SEQ ID NO: 19:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 797 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-700-013B-19

Query Match      32.2%; Score 1080.5; DB 2; Length 797;
Best Local Similarity 38.2%; Pred. No. 2e-93;
Matches 221; Conservative 106; Mismatches 213; Indels 39; Gaps 10.

OY 54 MSNIEFLMCSGSIARISVGLGNWVRPEFIAYONGGAGFLVPYIVILLVTKPYYIEBCVIGQ 113
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 WSKKIDFLSLWAGVAGVGNWVRPEFLYFQNGGAGFLPIYLMMLALAGLPIFFLEVSIGQ 253
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 114 FSSRSNVKWSISIPAMKGTGYAQAAGCGYILSYVVYIGCLLYLAMSFOATLPMAICQP 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 PASGPGVSWKRAIPALQCGIAMLILISVLAIYVNIICYFLFLFAFVSVLPMWGSCNN 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 174 EW-----ENCPSPD-PTL-----AASNNIT--NGTSSA-----QL 201
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 PWNTPEDCKDKTKLLDSCVISDHPKIQIKNSTFECTAPDNYTMVFTSQANKTFVSGSEE 373
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 202 YFLRTVLOQSDGIEGGLGAPIMYIVLCLFIAMLVFVGVANGVSKSGAAAYLALFPYV 261
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 YFKYFVLKISAGIEPR-GEIMPLALCLFLAMVIVYASLAKIKTSKQVYVFTATFPYV 432
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 262 MITEFITITILPGATDGLIFVYTPQWAKLLEGVYSAVTOVFESLYCTGPIMFSSYN 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 LVILIRSVTLPGAGAGIWEYFTEPKWEKLTATVWKDAATQIFFLSAAWGGLITLSYN 492
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 322 GFRHIVYDAMIVITLDTFTSPISGCTFFGLIGNLAYELNSVGVGAGTSLAIFSP 381
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 KFHNNCYDITLIVCTNSATISIFAGFVIFSVIGFMANERKNVNIENVADQ-GGGIAFPYR 551
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 382 DAIAKTFOPQLFSVLEFLIMSVLGIGSSVALLSTFNFLAMDAPRVPIVYNSAMT---CS 438
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 EALNRLPSPPWALIFFLMLTLTGLDTMFATIEYITVTSISDFEPKYLRTKHKVPFLIGCI 611
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 439 CGEILGLVYCPGGQGYILINDVHGCTFLVLCALISELAGVNIYGLNCLDIFPMIGX 498
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 612 CFFIMGPFMTIYGGIYMFQVLDVTAASTALVITLAFELGVISYVGLQFCDIEDIMIGF 671
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 499 KTGAVWRLCQWGTIVTAIMTVEFFVALLASNNLVFSDNYYVPRAGVSGYIMFLMGTEVP 558
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 672 QPNIFWAKCMAFVPTLITLFLICFSFYQWEPRTYG-SYRYPWMSVVLGMLMLACSVIWP 730
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 559 IGIGFSLYKRTGTSETIKKAFHSKPSMGPPSPREKRE 597
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 IMFVIMKH-LAPGRFIERLKLKVCSPQDPWGPELAAHGRGE 768

RESULT 5
US-09-191-468-122
/ Sequence 122, Application US/09191468A
/ Patent No. 6416975
/ GENERAL INFORMATION:
/ APPLICANT: Callagher, Michael J.
/ APPLICANT: Burgess, Loyd R.
/ APPLICANT: Bruggen, Kurt R.
/ TITLE OF INVENTION: Human Glycine Transporter Type 2
/ FILE REFERENCE: 123110501
/ CURRENT APPLICATION NUMBER: US/09/191,468A
/ NUMBER OF SEQ ID NOS: 124
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 122
/ LENGTH: 797
/ TYPE: PRT
/ ORGANISM: Human

```

Db 194 MSSKLDLFLISMVGAVGLGNWVRPFYLAQNQGAFLPIYIMMALALGLPIFFLEVLSTG 253

QY 174 EW-----ENCVPSD-PTL-----AASVNNIT--NGTSSA-----QL 201

Db 314 PWNTPECKDKTKLLDSCVISDHPKIQIKNSTFCMTAYPNVTMVNFTSQANKTFVSGSEE 373

202 YFLRTVLQOSDGI EGG LGAPIWYLVLCFLIAMLVFEVARGVKSSGKAYFLALFPYVV 261

Db 3/4 YEKYFVLKISAGIEP-GEIRWPLALCLELAWIVIVASLAKGIKTSKVVYFTATFPYV 432

202 M I L F I I I L P G A I D G I L F F V I P Q W A K L L E L G V M I S A V I Q V F E S L I V C I G P I M E S S Y N 321

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible][illegible]

```

553 FATRIBI.DSPWATTEEMITICIDMENAITEETVSTCEDEEVR BUHVREMI CCOCT 611
      :::  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
0b

```

439 CGELLGLVYCTPGGOYLLELVDHYGTELVLCFAISELAGVFWYGLENICLDTEFMIGK 498

Db 612 CFFIMGFPMTQGGIYMFQLDVYYAASYALVIAIFELVGISYVYGLQRCEDIEMMIGF 671

QY 499 KTGAYWRLCWGVITPAIMTVFFYALLASNLLVEGDNYVPTAGYVSGYLLMFLGMTFVP 558

Db 672 QPNI FKWCVAFVTP TILTFILCFSEFYQWEPMTYG-SYRYPNWSMVLGWLMLACSVIWP 730

559 LGIGFSLYKKRIGTGFSEILKKAHFHSKPSWGPSPRERRE 597

00 / 01 TIME INMM - LASTONE I EXENLVCSEFQEDWGEPE LAQHKGE / 006

## RESULT 7

```

; Sequence 120, Application US/09191468A
; Patent # 7,541,607

```

```

; GENERAL INFORMATION:
; APPLICANT: Callacher Michael T

```

APPLICANT: Brunden, Kurt R.

```

; TITLE OF INVENTION: human glycine transporter type 2
; FILE REFERENCE: 12311US01

```

CURRENT FILING DATE: 1998-11-12

```
; SOFTWARE: PatentIn Ver. 2.0
```

```

; LENGTH: 797
TYPE: PPT

```

US-09-191-468-120

Query Match 32.1%; Score 1076.5; DB 4; Length 797;

Matches	219;	Conservative	108;	Mismatches	220;	Indels	39;	Gaps	9
---------	------	--------------	------	------------	------	--------	-----	------	---

47 EPPERWWSNIEFLMSCIATSVGLGNWRFPFIAYQNGGAEVPIVILLVGKPVY 106

Db 187 EDKARGNWSKLDFTLSMVGAVGLGNVWRFPYLAQNGGAFLLPYLMLALAGLPFF 246

107 LECVLGQESSRNSVKWMSISPAMKGTGYAQAAGCGILSYVVICGLCLYLA MSFGATL 166

247 LEVSLGVFHSQGFVSVMNHLPALQGGIAMLISVLIAIYNVILCILFLFASFVSL 306

[illegible]

```

Db 307 PMGSCNNPMTPECKDKTKLLDSCVISHPKIQIKNSTFCMTAVPNTVNTSLANKT 366
Qy 197 --SSQLYFLRTVLOQSDIGEGGAPLWVLCFLTAMLVPGVAVGVSSGKAAYEL 254
Db 367 FVSGSEEFKRYFKISAGIEYP-GEIRMPALCLFLAMVIVASLAGITSGKVVYFT 425
Qy 255 ALFPVWMTFTFTIILPGATDGLFFVTPQMAKLELGYWASVAVTQVFSILVCTGPI 314
Db 426 ATFPYAVVILLINGVTLPGAGAGIWTFTPKMEKLDATVWKDAATQIFESLSAAMGL 485
Qy 315 IMSSYNGFRHNYRDAMVITTLDTFTSLSGCTIFGILNLAELNSEVDVAGAGTS 374
Db 486 ILLSYNKHNNCRDELIVTCTNSATSIIFAGFVFSVIGFMANERKNINENADQ-CPG 544
Qy 375 LAFISYPAIAKTQPOLFSVLFPLMSVIGSGSVALLSTFNLTAMAFRVPVYMSA 434
Db 545 IAFVYEPALTRLPSPFWAITFFLMLTLTGIDTFFASIEFTVVISISEPFRYLTRHPV 604
Qy 435 MT--CSCGFLGLVYCTPGGOYILELVHDYGGFELVFCASISLAGVFWIYGLNLCD 491
Db 605 FTLGCCICFIMGPFMTITGGIYMGQVDTYAAASALVIAIFELVGSIVYGLQRCED 664
Qy 492 IEFMLGKTGAIVRLCWCVITPAITVTFEYALLASNNLVGDNVYVTPAGVSGYLMF 551
Db 665 IEMMIGFOPNIFWKVCMAFVTPITLFTLFCESFYOMEPMTYG-SYRPNMSKVLGMLMA 723
Qy 552 LGMPFVPIGIFSLKRYTGFESETIKKAFKSPKSGRSPRERE 597
Db 724 CSYVWPIPMFVYIKM-LAPGRFIERKLKVCSPQDMPGFLAQRGE 768

```

## RESULT 8

US-08-700-013B-27

Sequence 27, Application US/08700013B

Patent No. 5919653

GENERAL INFORMATION:

APPLICANT: Albert, Vivian R.

APPLICANT: Kowalski, Leslie R.Z.

APPLICANT: Borden, Laurence A.

APPLICANT: McElvly, Jeffrey F.

TITLE OF INVENTION: Human Glycine Transporter

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESS: Dechert Price &amp; Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,013B

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 317743-108

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 799 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-700-013B-27

Query Match

Best Local Similarity 32.1%, Score 1076.5; DB 2; Length 799;

Matches 224; Conservative 107; Mismatches 228; Indels 39; Gaps 10;

```

Qy 35 ALDNDIDDDLEAPPERBERVWNNIEFLMSCTASVIGVNWRRPFVAVONGGAFVAVY 94
Db 177 ATTIPEDGODENKARGMSSKIDFLISMGVAVGNWRRPFVAVLAFONGGAFVAVY 236
Qy 95 IYLLVGRPVYLLCEVIGGOFSSRNSVYKWSISPAKMGTYGAQAAGCYLLSYVYIGLC 154
Db 237 MMLALAGLPIFLESLGQFASQGVSWKAPALOGGICMLLISVLAIYVAVIICYT 296
Qy 155 LYIYLAASQATLPALICOPEN-----ENCYPSD-PLI-----AASVNI 192
Db 297 LEYLFASVSVLPMGSCNNPMTPECKDKTKLLDSCVIGHPKIQIKNSTFCMTAVPVL 356
Qy 193 T--NGTSSA-----OLYFLRTVLOQSDIGEGGAPLWVLCFLTAMLVPGVAVAR 242
Db 357 TMAVNTSQANKTPVSGSEEFKRYFKISAGIEYP-GEIRMPLEPCFLFAMVIVASLAK 415
Qy 243 GYKSSGKAAYFLALFPYVYMTLFTIILPGATDGLFFVTPQMAKLELGYWASVATQ 302
Db 416 GIKTSGKVYFLAFPPYVAVVILLIRGVTLPGAGAGIWTFTPKMEKLTDAVWKDAATQ 475
Qy 303 VFESLTVCTGPIHMSYNGFRHNYRDAMVITTLDTFTSLSGCTIFGILNLAELNS 362
Db 476 IFFSLAAMGGILTLISSYNNKFNHNCYRDTLVCTNSATSIIFAGFVFSVIGFMANERKV 535
Qy 363 EYGDVAVAGTSLAFISYPAIAKTQPOLFSVLFPLMSVIGSGSVALLSTFNLTAM 422
Db 536 NIENAVDQ-GGCIAPVYFPAITRLPSPFWAITFFLMLTLTGIDTFFASIEFTVVIS 594
Qy 423 AFRVPVYVMSAMT--CSCGFLGLVYCTPGGOYILELVHDYGGFELVFCASISLAGV 479
Db 595 EEPKYLRTHKPVPFTLGCCICFIMGPFMTITGGIYMGQVDTYAAASALVIAIFELVGI 654
Qy 480 FWYGLNLCDIEEMLGKKTGAIVRLCWCVITPAITVTFEYALLASNNLVGDNVYV 539
Db 655 STYVGLORCEDIEEMIGFOPNIFWKVCMAFVTPITLFTLFCESFYOMEPMTYG-SYR 713
Qy 540 TAGVYGYLMLFLGTFTVPIGIFSLYRYGTGFESETIKKAFKSPKSGRSPRERE 597
Db 714 NMSVGLNMLLACSVIYIPIMVYIKM-LAPGRFIERKLKVCSPQDMPGFLAQRGE 770

```

## RESULT 9

US-07-879-617A-8

Sequence 8, Application us/07879617A

Patent No. 5580775

GENERAL INFORMATION:

APPLICANT: Tremaine Jr., Robert T.

APPLICANT: Caron, Marc G.

APPLICANT: Blakely, Randy D.

TITLE OF INVENTION: A High Affinity L-Proline Transporter

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESS: Kilpatrick &amp; Cody

STREET: 1100 Peachtree Street, Suite 2800

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.

ZIP: 30309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/879,617A

```

1 FEATURE:
2 NAME/KEY: Region
3 LOCATION: 44..45
4 OTHER INFORMATION:
5 OTHER INFORMATION:
6 FEATURE:
7 NAME/KEY: Region
8 LOCATION: 71..72
9 OTHER INFORMATION:
10 OTHER INFORMATION:
11 FEATURE:
12 NAME/KEY: Region
13 LOCATION: 269..270
14 OTHER INFORMATION:
15 OTHER INFORMATION:
16 FEATURE:
17 NAME/KEY: Region
18 LOCATION: 630..631
19 OTHER INFORMATION:
20 OTHER INFORMATION:
21 FEATURE:
22 NAME/KEY: Region
23 LOCATION: 106..127
24 OTHER INFORMATION:
25 OTHER INFORMATION:
26 FEATURE:
27 NAME/KEY: Region
28 LOCATION: 212..213
29 OTHER INFORMATION:
30 OTHER INFORMATION:
31 US-07-879-617A-8

```

Query Match	30.7%	Score 1030.5	DB 1	Length 667
Best Local Similarity	38.7%	Pred. No. 8.3e-892		
Matches	234	Conservative	94	Mismatches 235; Indels 21; Gaps 13
QY	11	ESSEPKMEPKRSSQISLT--PPANNKRALDNID--DNDLAE--PPERMMVSNNEFLMSCIA	66	
Db	23	ESEQENCKMKIKQEAHLRKPVTPLDLMPSPDGGVDVLDVFAADRGNTGKDLFLLSCIG	82	
QY	67	TSVIGNWRREPPFLAYONGGAFVLPYIVILLNCKRPYYLLECVLGQSSRSRVVWS	126	
Db	83	YCVGIGNWRRPYRYITNNGGAFVLPYFLMLAICIPLEFELSGQSSISGLPRAWKIS	142	
QY	127	PAMKGTGYAAGCGYILSYVVVIGCLLYLWAMSFQATLPMALCPBM--ENCY----	180	
Db	143	PLFGAGAGAMLLIIGLVAIYYNMIIAYVLFLASLTNSLPMHEGCMNMNERCLEHNGP	202	
QY	181	SDPLLASVNNITNGTSSAQLXELRTYL--QOSDIEGGLCAPRYLVLYLCLFIAMLYPFG	238	
Db	203	KDNGCALPL--NLSTSVSPSEETWSRYVLIHIGSQSI--GRPEIRINLCLILLAAVIVFL	260	
QY	239	VWARGVSSGKAAYFLALFPYVMITLFTITLIIIPGARDGILFEVTPOMAKLLEGVWS	298	
Db	261	CILGKVGSSGKVYVFTATLFPYLLILMLLVRGVTLPGAMKGIQFYLTPQFHHLLSKWVIE	320	
QY	299	AVNQVEFSLYMCQPIINNESSYNGFRNHRYDANIVTLTDLPFTSLSCITFGILIGNLAY	358	
Db	321	AAIDIFSLSGFGGLLTFEASYNFHOHYTDFTIVLIGNAITSLLAFAFLFVSGVMSQ	380	
QY	359	ELNSEVGDVVGAGTSLAFTSPDAIKTFQOPQFSLVFELFMSVGLIGSSVALISTENT	418	
Db	381	ELGVPV--DOYAKAGPGLAFVITYPQAMMLPLSPFMSLFFPMLLTIGDSQFALEIYVT	439	
QY	419	LADAFP---RVLPYVMSAMTSCSGFLDLGLVYCPGGQYILLEVYDHGGTFLVFCALISE	475	
Db	440	AVTDEFPYLRPKRAVESGLICVAMVYLMGLLTITDGGYMWLVLLDDVSASFGVAVVYTT	499	
QY	476	LAGFVFIYGLLENCLDIEFMLGKRTGAYWRLCMQVITPAIMTGVFFALLASNMLVPCDN	535	
Db	500	CLAVTRYGYQIQRCDIHMMLGFRPGLYFRACWLFLESPATLLALVYSIVYQSEVQ--S	558	
QY	536	YVYPTAGYVGYLMLFLGNTFVPIGIGFSLKYKRTGTFSESTIKAFHSPKSGRSPRER	595	

Db 559 YRRPAPAEILGILMGILSCLMIPACMLAVLR-EEGSLMERLQQAAPRAIDMGF-SLEEN 616

QY 596 REMM 599

Db 617 RTGM 620

```

RESULT 10
US-08-753-985-8
: Sequence 8, Application US/08753985
: Patent No. 5759788
:
: GENERAL INFORMATION:
: APPLICANT: Fremeanu Jr., Robert T.
: APPLICANT: Caron, Marc G.
: APPLICANT: Blakely, Randy D.
: TITLE OF INVENTION: A High Affinity L-Proline Transporter
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: U.S.
: ZIP: 30309
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/753,985
: FILING DATE: 03-Dec-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/879617
: FILING DATE: 01-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Padst, Patrea L.
: REGISTRATION NUMBER: 31,284
: REFERENCE/DOCKET NUMBER: EMU109
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6508
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 667 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: Rattus
: TISSUE TYPE: Brain
: IMMEDIATE SOURCE:
: LIBRARY: rat forebrain cdna library
: CLONE: rTB2-2-20
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 76..95
: OTHER INFORMATION:
: OTHER INFORMATION: /note= "Membrane-spanning domain"
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 103..127
: OTHER INFORMATION: /note= "Membrane-spanning domain"
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 147..167
: OTHER INFORMATION: /note= "Membrane-spanning domain"
: FEATURE:
: NAME/KEY: Domain

```

LOCATION: 247..266	/note=	"Membrane-spanning domain"
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: Domain		
LOCATION: 272..294		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 321..341		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 404..427		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 454..473		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 487..509		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 530..549		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Domain		
LOCATION: 567..589		
OTHER INFORMATION:	/note=	"Membrane-spanning domain"
FEATURE:		
NAME/KEY: Region		
LOCATION: 44..45		
OTHER INFORMATION:	/note=	"protein kinase C
OTHER INFORMATION:		phosphorylation site"
FEATURE:		
NAME/KEY: Region		
LOCATION: 71..72		
OTHER INFORMATION:	/note=	"cAMP-dependent protein
OTHER INFORMATION:		kinase phosphorylation site"
FEATURE:		
NAME/KEY: Region		
LOCATION: 268..270		
OTHER INFORMATION:	/note=	"protein kinase C
OTHER INFORMATION:		phosphorylation site"
FEATURE:		
NAME/KEY: Region		
LOCATION: 630..631		
OTHER INFORMATION:	/note=	"protein kinase C
OTHER INFORMATION:		phosphorylation site"
FEATURE:		
NAME/KEY: Region		
LOCATION: 106..127		
OTHER INFORMATION:	/note=	"leucine zipper motif"
FEATURE:		
NAME/KEY: Region		
LOCATION: 212..213		
OTHER INFORMATION:	/note=	"N-linked glycosylation
OTHER INFORMATION:		site"

	Query Match	30.7%	Score 1030.5;	DB 1;	Length 667;
	Best Local Similarity:	38.7%	Pred No. 8.3e-89;		
	Matches 234;	Conservative 94;	Mismatches 255;	Indels 21;	Gaps 13;
OY	11 ESSEPKMKRSSQSLP-PANNKAALDNID--PTDLEAF-PPERWMSNNIEFLMSCIA	66			
	:::				
DB	23 ESDGNCMKKLTQEENHLRKPYTPDLPLMPDSOGDDVDLVAADAGNGATGKIDFLSTIG	82			

QY 67 TSVGLGNVRRPFIATYONGGAFIIVPVIVLLGVKPYVLECYLGQSSHNKYSKVSIS 126  
 Db 83 YCVGLGNVRRPFIATYONGGAFIIVPVIVLLGVKPYVLECYLGQSSHNKYSKVSIS 142  
 QY 127 PAMGCTGAAAGCGIILSYIVTCGICLYLANSFOATLPMAICQEBW--ENCV----P 180  
 Db 143 PLEFGAAGAAMLIIVGLAVIYNNMTAVLEFLFASLTSNLPWEHCGMMWNERCLEHGP 202  
 QY 181 SDPLTAAVNNITNGTSSAQLYFRTVL--QOSDIGGAGAPIMYLVLCFIMLVWFG 238  
 Db 203 KDGNGALPL-NLSTVSPSEYKSRVYLHIOGSOGI-GRGEIYNNICLICLLMWIVFL 260  
 QY 239 VVARGVSSGKAAYELALFPYVNTLEFITTIIIPGATDGLFVTPQMAKLELGYWS 298  
 Db 261 CILKGVSSGKVVYFATFAPYLLMLLVGVTLPGAMKGIQVFLTFQFHLLSSKWIE 320  
 QY 299 AVTQVFEFLVCGPIIMFSSYNGERNITRDAMIVTLDITFSLGCTIFGLGNLAY 358  
 Db 321 AALQIFSLGVGEGGLTFASYNTHONITRDTFIVTLGNATISILAGFALFVLYGMSQ 380  
 QY 359 ELNSEVDVYAGGTSIAFTSYPAIAKTFOPQLFSLVFLMMSVLGSSVALLSTENT 418  
 Db 381 ELGVYV-DQYAKAGPGLAVIYPCAMTMLPLSPWSPLEFFMLTLTLDQFALEIYVT 439  
 QY 419 LADAPF--RVPVYNSAMTCSGIFLLGVYCTPGQVILELVDHXGFFVLVCAISE 475  
 Db 440 AVTDEFYVLRPKKAVESGLICVAVYLMGLILTTDGMVYLVLLDDYSASFGLAVVIT 499  
 QY 476 LAGVFWYGLNCLDIEFMIGKKIGAYWRLCWGVITPAIMTYVFFALLASNNLVGDN 535  
 Db 500 CLAVTRYGIQRCDRIHMLGFRGGLYFRACWLELSPALLALVSIYKOPSEYG-S 558  
 QY 536 YVYPTAGVSGYILFLGMTFVPLIGFSLYKRTGTSETIKAKFSSKMGPSRSPR 595  
 Db 559 YREPAMAILGILMGLISCLMIPASMLVAVLR-EGSLMRLQASPAIDWGP-SLEEN 616  
 QY 596 REMM 599  
 Db 617 RTGM 620

TELEFAX: 404-815-6555  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 635 amino acids  
 TYPE: AMINO ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: Rattus  
 TISSUE TYPE: Brain - Proline Transporter  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 46..65  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 72..97  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 117..137  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 217..236  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 243..264  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 322..345  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 373..397  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 424..443  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 456..479  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 500..519  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 536..559  
 OTHER INFORMATION: /note= "Proposed transmembrane domain."

RESULT 11  
 US-07-879-617A-9  
 Sequence 9, Application US/07879611A  
 Patent No. 5580775  
 GENERAL INFORMATION:  
 APPLICANT: Freneau Jr., Robert T  
 APPLICANT: Caron, Marc G  
 APPLICANT: Blakely, Randy D  
 TITLE OF INVENTION: A High Affinity L-Proline Transporter  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kilpatrick & Cody  
 STREET: 1100 Peachtree Street, Suite 2800  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: U.S.  
 ZIP: 30309  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/879,617A  
 FILING DATE: 19920501  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pabst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE/DOCKET NUMBER: EMU03  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 404-815-6508

OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 76..77  
OTHER INFORMATION: /note= "Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 83..84  
OTHER INFORMATION: /note= "Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 90..91  
OTHER INFORMATION: /note= "Leucine zipper motif"  
FEATURE:  
NAME/KEY: Active-site  
LOCATION: 97..98  
OTHER INFORMATION: /note= "Leucine zipper motif"  
US-07-819-617A-9

Query Match 30.3%; Score 1017; DB 1; Length 635;  
Best Local Similarity 39.4%; Pred. No. 1.5e-87;  
Matches 225; Conservative 91; Mismatches 237; Indels 18; Gaps 11;

QY 41 DTDLAE-PPERYVSNNIEFLMSCIATSVGLNVRFPFLAYONGGAFILPVYVLL 99  
D 26 DVDVDVDFADRGWMTKIDFLSLCIGYCVGLGNVWFRPYRAYTNGGAFILPVYVLL 85  
QY 100 VGRPVYLLCEVLFSSRNYSKVYVSPAMKGTGACAGGCTLSYVVCGLIYVLA 159  
D 86 CGIPLEFLSLGOFSSGLPLAVWKISPLFEGAGAMLLYGLAITYNMIAVLYL 145  
QY 160 MSFOATLPMAICPEW--ENCV---PSDPLASVNNITNGTSSAOLYFLRVL--QOS 211  
D 146 ASLYSNLPWEHCGMMWTERCLEHRGPKDNGALPL-NTSTVSPSEYVRYVLIHGS 204  
QY 212 DGTGGGAPRIWYVLCFLTAMLVFGVAVGSSGKAYFLALFPYVMTITFTTII 271  
D 205 QGI-GRPELRWNLCLILLAWIVFLCIIKGVKSSGRVYFTATFPYLLIMLVKGT 263  
QY 272 LPGATDGLFEVFPOMAKLELVGYSAVTQVFSLVNCTGPIIMFSSYNGFRNIRDA 331  
D 264 LPGAOKIOFYLLPQFHLHLLSKYVIRAAIOIFSLVGRGGLTFPASYNTHONIRDT 323  
QY 332 WIVTTLDFTSFLSGCTIFGLNMAELSEVGVAGGTSIAFTSYDPAIAKTEPOQ 391  
D 324 FIYTLGNATISIIAGFAIFSVLYGMSQELGVPV-DYAKAGPGLAIFYIPQAMMLPSP 382  
QY 392 LFSVLFPLMSVIGISGVSALLSTFTNLADAPP--RVPTVYSAMTCCSGFILGLVYC 448  
D 383 FWSFLFEMLLTGLDQFAFLFETIVAVTDEFPYLRPKKAVFSGLICVAMYIMGLILT 442  
QY 449 TPGGOYILELVHDHGFVLFCAISELAGFWITGLENCLDIEFMIGKGTAYVRLCW 508  
D 443 TDGGMVYLIDYASFGMLVYVITICLATVRYVGIQRCDRIHMLGEPGILYFRACW 502  
QY 509 GVTPALMTVFYFALLASNNLVFGDNYVPTAGVSGYLMFLGTFVPYIGISFLYKY 568  
D 503 LFLSPATILALLVYSIKYQPSYRG-SYRPPAMAEILGILMGLSCIMIPAGMLVAVLR- 560  
QY 569 RTGFFSTIKKAFHSKSPGPRPREREM 599  
D 561 EESGLMERLDOASRPADWGP-SLEENRTGN 590

RESULT 12  
US-08-753-985-9  
Sequence 9, Application US/08753985  
Patent No. 5759788  
GENERAL INFORMATION:  
APPLICANT: Fremeau Jr., Robert T.  
APPLICANT: Caron, Marc G.  
APPLICANT: Blakely, Randy D.  
TITLE OF INVENTION: A High Affinity L-Proline Transporter

NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/753,985  
FILING DATE: 03-DEC-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/879617  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Padst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Rattus  
TISSUE TYPE: Brain - Proline Transporter  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 46..65  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 72..97  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 117..137  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 217..236  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 243..264  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 291..311  
OTHER INFORMATION: /note= "Proposed transmembrane  
OTHER INFORMATION: domain."  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 322..345

Query Match	30.3%	Score	1017	DB 1	Length	635
Best Local Similarity	39.4%	Pred	No. 1.5e-87			
Matches 225	Conservative	91	Mismatches 237	Indels	18	Gaps 11

```

OY      392 LEVSLEFLMMSVGLGSSVALLSPNTLAMAAP---RVPPVYMSAMKSCGFLGLGYC 448
Db      383 FWSLFEFEMLLTGLGDSQFAFLFETTVIAVTEEFPPYLPKRAVSGLLCVANYLMLGLLT 442
OY      449 TPGQGYILELDVHYGFTLVFLCAISELAGVFIAGLENLCIDIEFMIGKRTGAYWRICW 508
Db      443 TDGGWYMWLVLLDDYSASRGLMWVYITTCLATVTRYGICRPFCDHIMGLGFKRGLFRAWC 502
OY      509 GVTPPAITVTFYFFALLASNNLVGDNVYVPLAGVSGTLMFLGMPVPIGIGSLYK 568
Db      503 LFLSATLALLVYSIVKYQPSXG-SYRFPAMAEILGILMLGLLSCLMIPAGMLVAVLR- 560
OY      569 RTGFSEETIKKAFHSKPSGCRSPREREM 599
Db      561 EEGSIMERLQOASRAIDMGP-SLEENRTGM 590

RESULT 13
US-08-301-722A-5
Sequence 5, Application US/08301722A
Patent No. 5756307
GENERAL INFORMATION:
APPLICANT: Uhl, George R.
APPLICANT: Vandenberg, David
APPLICANT: Persico, Antonio
TITLE OF INVENTION: SEQUENCE OF HUMAN DOPAMINE TRANSPORTER
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESS: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/301,722A
FILING DATE: 07-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1173-406P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 599 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
FEATURE:
NAME/KEY: Protein
LOCATION: 1..599
OTHER INFORMATION: /note= "Hgbat sequence, see Fig.
OTHER INFORMATION: 5"
US-08-301-722A-5

Query Match 30.1%; Score 1011; DB 1; Length 599;
Best Local Similarity 35.4%; Pred. No. 4.9e-87;
Matches 213; Conservative 118; Mismatches 229; Indels 42; Gaps 13;

```



Db 10 DGOI-----STEVSAPVANDKPKTLVVKVOKKADLPDRDTWKGKRPDL 54  
Qy 62 MSCATSVGLGNVWRFPIAYONGGAFVLPVYVLLVGRPVYVLECVLGQFSSRNSVK 121  
Db 55 MSCGYAIGLGNVWRFPIAYONGGAFVLPVYVLLVGRPVYVLECVLGQFSSRNSVK 114  
Qy 122 VMSISPAKGTGYAQAACGYILSYVYVIGCLCYLLAMSFQATLPMAICQPEW--ENCV 179  
Db 115 VMLKAPMKGVGLAAVLSFVNLNTIYIYIISMAIYLYLNSFTTLPMKQCNPMWTDKCF 174  
Qy 180 PSDPTLAASVNNITNGTSSAOLYFRLYLOOSDIEGGLGAPVWLVLCPLIAMLVPGV 239  
Db 175 SN-----YSLVNTNMTSAVVEFERNHOMTDGDKP-GQIRWPLATTLIAWLVYFC 228  
Qy 240 VARGVSSGKAAYFLATLPVYVMTLFTTTLILPGATGGLFVVPQMAKLELGWYSA 299  
Db 229 IMKGVGWTGVYVSATVPYIMLILFPRGYTLPRQAKGILFYITPNRKLSDSEVWIDA 288  
Qy 300 VTQVFFSLVCTGPIFMSYNGFRHNYRDAMVITLDTFTSLGCTIGILGNLAYE 359  
Db 289 ATQIFFSYGLGSLALAGSYNSFNHNYRDSIIVCCINSCTSMFAGVIFSIYGFMAHV 348  
Qy 360 LNSEGVYVAGGSLAFISYDPAIAKTFQPOLSVLEFLMMSVIGISSVALLSTFNTL 419  
Db 349 TKRSIAD-VAAAGGLAFVLAPEAVTQPLISPLMALIFFSMLMLGIDSOCTYEGFTTA 407  
Qy 420 AMDAFPRV-----PYVMSAMTSCGFLGLVYCTPGQYILEVDHYGFTL-VLFCAS 474  
Db 408 LVDEYPRLLRNRRELFLAAV-CIISYLGISNITOGGIYVFKLPFYASASMSLFLVVF 466  
Qy 475 ELAGVFWIYGLNCLDIEFMLGKKTGAYWRLCWGVTPAIMTYVFFYALLASNNLVFGD 534  
Db 467 ECVSISWYGVNRFYDNIQENWVGRPCIMWKLCSFPIIYAGVIFSAVQMPPLWNG- 525  
Qy 535 NYVYPTAGVSYGLMFLGMTFVPIGIGFLYKXRT--GTFSEITKKAFFHS-----KRW 587  
Db 526 NYVEPKMGQGVGMALSSMVLIP---GYMAVYMLLAKGSLKORQVWQPSSEDTVRPN 582  
Qy 588 GP 589  
Db 583 GP 584

RESULT 14  
US-08-295-814E-11  
Sequence 11, Application US/08295814E  
Patent No. 5658786  
GENERAL INFORMATION:  
APPLICANT: Smith, Kelli E.  
APPLICANT: Borden, Laurence A.  
APPLICANT: Hartig, Paul R.  
APPLICANT: Weinschank, Richard L.  
TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA  
TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/295,814E  
FILING DATE: DECEMBER 19, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John

REGISTRATION NUMBER: 28, 678  
REFERENCE/DOCKET NUMBER: 40558-B-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 599 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-295-814E-11

Query Match 30.1%; Score 2008; DB 1; Length 599;  
Best Local Similarity 34.8%; Pred. No. 9,5e-87;  
Matches 211; Conservative 122; Mismatches 231; Indels 42; Gaps 13;

Qy 3 DGOVNGGFESSEPKMKPRSSQIS-LPPANNKALDNIDPDLEAEPERWVSNIEFL 61  
Db 10 DGOI-----STEVSAPVANDKPKTLVVKVOKKADLPDRDTWKGKRPDL 54  
Qy 62 MSCATSVGLGNVWRFPIAYONGGAFVLPVYVLLVGRPVYVLECVLGQFSSRNSVK 121  
Db 55 MSCGYAIGLGNVWRFPIAYONGGAFVLPVYVLLVGRPVYVLECVLGQFSSRNSVK 114  
Qy 122 VMSISPAKGTGYAQAACGYILSYVYVIGCLCYLLAMSFQATLPMAICQPEW--ENCV 179  
Db 115 VMLKAPMKGVGLAAVLSFVNLNTIYIYIISMAIYLYLNSFTTLPMKQCNPMWTDKCF 174  
Qy 180 PSDPTLAASVNNITNGTSSAOLYFRLYLOOSDIEGGLGAPVWLVLCPLIAMLVPGV 239  
Db 175 SN-----YSLVNTNMTSAVVEFERNHOMTDGDKP-GQIRWPLATTLIAWLVYFC 228  
Qy 240 VARGVSSGKAAYFLATLPVYVMTLFTTTLILPGATGGLFVVPQMAKLELGWYSA 299  
Db 229 IMKGVGWTGVYVSATVPYIMLILFPRGYTLPRQAKGILFYITPNRKLSDSEVWIDA 288  
Qy 300 VTQVFFSLVCTGPIFMSYNGFRHNYRDAMVITLDTFTSLGCTIGILGNLAYE 359  
Db 289 ATQIFFSYGLGSLALAGSYNSFNHNYRDSIIVCCINSCTSMFAGVIFSIYGFMAHV 348  
Qy 360 LNSEGVYVAGGSLAFISYDPAIAKTFQPOLSVLEFLMMSVIGISSVALLSTFNTL 419  
Db 349 TKRSIAD-VAAAGGLAFVLAPEAVTQPLISPLMALIFFSMLMLGIDSOCTYEGFTTA 407  
Qy 420 AMDAFPRV-----PYVMSAMTSCGFLGLVYCTPGQYILEVDHYGFTL-VLFCAS 474  
Db 408 LVDEYPRLLRNRRELFLAAV-CIISYLGISNITOGGIYVFKLPFYASASMSLFLVVF 466  
Qy 475 ELAGVFWIYGLNCLDIEFMLGKKTGAYWRLCWGVTPAIMTYVFFYALLASNNLVFGD 534  
Db 467 ECVSISWYGVNRFYDNIQENWVGRPCIMWKLCSFPIIYAGVIFSAVQMPPLWNG- 525  
Qy 535 NYVYPTAGVSYGLMFLGMTFVPIGIGFLYKXRT--GTFSEITKKAFFHS-----KRW 587  
Db 526 NYVEPKMGQGVGMALSSMVLIP---GYMAVYMLLAKGSLKORQVWQPSSEDTVRPN 582  
Qy 588 GPRSPR 593  
Db 583 GPEQPO 588

RESULT 15  
US-08-240-783B-4  
Sequence 4, Application US/08240783B  
Patent No. 5756348  
GENERAL INFORMATION:  
APPLICANT: Smith, Kelli  
APPLICANT: Borden, Laurence A.  
APPLICANT: Hartig, Paul R.  
APPLICANT: Weinschank, Richard L.

TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.14  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,783B  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1794/39875-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212)391-0525  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 599 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: N  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: RAT GABA TRANSPORTER (GAT-1)  
US-08-240-783B-4

Query Match 30.1%; Score 1008; DB 1; Length 599;

Best Local Similarity 34.8%; Prd No. 9.5e-87;  
Matches 211; Conservative 122; Mismatches 231; Indels 42; Gaps 13;

QY 3 DGOVNGGESSEPKMEKRSQIS LPPANKKALDNIDDTLEAEPPERWWSNNIEFL 61  
DB 10 DGOI-----STENSIAPASDKPKTLVVKQKAGDLPDRDTWKGREDFL 54  
QY 62 MSCATISVGLGNWRFPIAYONGGAFVPIYVILLVKGPRVYEECVLGQSSRSNYSK 121  
DB 55 MSCGVYAIGLGNWRFPIYLCGKNGGAFLLPYFTLLFAGVPLEFLLECSLGQYISIGLG 114  
QY 122 VMSISPAKGTGYAQAAGCYILSYVYVIGLCIYLAWSFOATLPWAICPEW--ENCV 179  
DB 115 VMKLAPEKGVGLAANLFSWIMIVYIISMALYIYNSFTITLPPKQCDNPWNTDRCF 174  
QY 180 PSDPTLAASVNNITNGTSSAQLFTRTVLOQSDGIEGGLGAPIWYLVLCFIAMLVFV 239  
DB 175 SN-----YSLVNTNMTSAVVEWIRNMHOMTDGLDKP-QQIRWPLAITLAIAVLYFC 228  
QY 240 VARGVKSAGAAAYELAFVYVNTLFTITLIPGATDGLFFVTPQMAKLLIEGVYXA 299  
DB 229 IMGVGWTGVVFSAYPIYIMILFFRGVTLPGAREGILFYITPNFRKLSDEVWLDA 288  
QY 300 VTQVFFSLVCTGPIIMFSSYNFHNHYRDAMIVTLDFTFSLGCTFGILGNLAYE 359  
DB 289 ATQIFEFYGLGLSLILASYNFNNVYRDSITVCCINCSIMFAGFVIFSIYGFMAHV 348  
QY 360 LNSEVGDVAGGTSLAFISYPAIAKTFQPLFSVLFELMSVLIIGSSVALLSTPNTL 419  
DB 349 TKRSIAD-VAASGGLAFIAYTOLPISPLAAILFESMLMLGLIDSOFCFVEGFTTA 407  
QY 420 AMDAFPRV-----PIVYMSAMTCGFLGLGVYCTPGQYILLEVDHIGGTFLL-VLPCAIS 474  
DB 408 LVDEYPRLLRNRLRLTAAY-CIVSYLIGLSNITGGIYVFKLEDYYSASGMSLLFLVFF 466

QY 475 ELAGVEWVIGLENCLDIEEMICKKTGAVWRLQMGVITPAIMTTFEYALLASNNLYFGD 534  
DB 467 ECYSISWFYGVNRRPDNIQEMVGRPCITWKKLMSFPTPLIVAGVFLESVAVOMTPLTMG- 525  
QY 535 NYVYPTAGVYGYLMLFLGNTFVPIGIFSLYKRT--GTFSETIKKAFHS-----KPSW 587  
DB 526 SYVEPKWGGGVGNMLMALSSVLLP---GYMAWYFLTLKSLKQRLQVMIQPSEDIYVRPN 582  
QY 588 GPRSPR 593  
DB 583 GPEQPQ 588

Search completed: July 21, 2003, 09:28:09  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:27:55 ; Search time 57 Seconds  
(without alignments)  
1318.858 Million cell updates/sec

Title: US-09-991-458-2

Perfect score: 3354

Sequence: 1 MNDQVNGCGESSEPEKMKER.....TSVKHLWYSITGAYRRNIN 633

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published\_Applications\_AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep: \*  
2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep: \*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep: \*  
6: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep: \*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep: \*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep: \*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep: \*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep: \*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep: \*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep: \*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep: \*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep: \*  
15: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep: \*  
16: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep: \*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1082.5	32.3	797	US-09-795-232-2	Sequence 2, Appli
2	1025.5	30.6	624	US-09-795-693-24	Sequence 24, Appli
3	1025.5	30.6	624	US-10-156-239-24	Sequence 24, Appli
4	1025.5	30.6	624	US-10-156-239-24	Sequence 24, Appli
5	991	29.5	630	US-09-843-598-11	Sequence 11, Appli
6	971	29.0	622	US-09-843-598-11	Sequence 11, Appli
7	953.5	28.4	587	US-09-815-923-4	Sequence 4, Appli
8	932.5	27.8	514	US-09-919-039-378	Sequence 378, App
9	929.5	27.7	597	US-09-815-923-14	Sequence 14, Appli
10	882	26.3	727	US-09-815-923-10	Sequence 10, Appli
11	869	25.9	556	US-09-815-923-6	Sequence 6, Appli
12	852.5	25.4	671	US-09-843-598-5	Sequence 5, Appli
13	844.5	25.2	671	US-09-843-598-7	Sequence 7, Appli
14	833	24.8	676	US-09-815-923-12	Sequence 12, Appli
15	790	23.6	730	US-09-741-149-2	Sequence 2, Appli
16	790	23.6	730	US-09-795-693-5	Sequence 5, Appli

17	790	23.6	730	US-10-156-239-5	Sequence 5, Appli
18	790	23.5	729	US-10-199-485-5	Sequence 5, Appli
19	787.5	23.6	730	US-09-741-149-4	Sequence 4, Appli
20	777.5	23.2	599	US-09-861-846-4	Sequence 4, Appli
21	768.5	22.9	727	US-09-923-444A-2	Sequence 2, Appli
22	767.5	22.9	610	US-09-861-846-2	Sequence 2, Appli
23	755	22.5	437	US-09-818-656A-4	Sequence 4, Appli
24	755	22.5	437	US-10-216-441-4	Sequence 4, Appli
25	755	22.5	459	US-09-818-656A-2	Sequence 2, Appli
26	755	22.5	459	US-10-216-441-2	Sequence 2, Appli
27	594.5	17.7	421	US-09-843-598-9	Sequence 9, Appli
28	469	14.0	579	US-09-738-628-4648	Sequence 4648, Ap
29	321.5	9.6	449	US-09-769-787-110	Sequence 110, App
30	274	8.2	224	US-09-843-598-6	Sequence 6, Appli
31	133	4.0	526	US-10-156-761-14237	Sequence 14237, A
32	128.5	3.8	305	US-10-183-116-27	Sequence 87, Appli
33	128	3.8	196	US-09-989-442-87	Sequence 4363, A
34	127	3.8	52	US-09-864-761-43363	Sequence 329, App
35	120.5	3.6	111	US-10-080-170-329	Sequence 4, Appli
36	114.5	3.4	619	US-09-741-153-4	Sequence 11835, A
37	113.5	3.4	755	US-10-156-761-11835	Sequence 11824, A
38	113	3.4	658	US-09-815-242-11824	Sequence 39755, A
39	112.5	3.4	84	US-09-864-761-39755	Sequence 4718, Ap
40	112.5	3.4	400	US-09-738-626-4718	Sequence 12636, A
41	112	3.3	497	US-09-815-242-12626	Sequence 11936, A
42	111.5	3.3	471	US-09-815-242-11936	Sequence 10014, A
43	111.5	3.3	476	US-09-815-242-10014	Sequence 4890, Ap
44	110.5	3.3	431	US-09-815-242-4890	Sequence 50, Appli
45	110	3.3	609	US-09-801-368-50	

## ALIGNMENTS

## RESULT 1

US-09-795-232-2

Sequence 2, Application US/09795232

Patent No. US20010012627A1

GENERAL INFORMATION:

APPLICANT: Anthony M. Brown

APPLICANT: Conrad Gerald Chapman

APPLICANT: Israel Simon Gloger

APPLICANT: Joanne Rachel Evans

APPLICANT: William Cairns

APPLICANT: Hugh Jonathan Herdon

TITLE OF INVENTION: NOVEL COMPOUNDS

FILE REFERENCE: GP-30176-D1

CURRENT APPLICATION NUMBER: US/09/795,232

CURRENT FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: 09/182,728

PRIOR FILING DATE: 1998-10-29

PRIOR APPLICATION NUMBER: 9818890.7

PRIOR FILING DATE: 1998-08-28

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 797

TYPE: PRT

ORGANISM: HOMO SAPIENS

US-09-795-232-2

Query Match 32.3%; Score 1082.5; DB 10; Length 797;

Best Local Similarity 38.3%; Pred. No. 1.6e-89;

Matches 222; Conservative 105; Mismatches 213; Indels 39; Gaps 10;

QY	54	WSNNIEPLMSCIATSVGLGNVWRPFVIYONGGAFVPIYVLLVGVKPYIECVGQ	113
DB	194	WSSKIDFLISMVGVAVGLGNVWRPFVLAFOGNGAFILPYLMMLALAPLPIFFLEVSIGQ	253
OY	114	FSSRSYKVMYISPMKGTGAQAAGCGYIISYVVICGLCTIYIAMSFOATLPAICOP	173
DB	254	FASQGVSVWKAIPALOGGCIAMLIISVLAITYYVNIICYTLFIYFASFVSILPWGSCNN	313

174 EM-----ENCPSH-----AASVNNIR--NGTSSA-----QL 201  
314 PANNTECKKTKTLLDSCVISHKIQIKNSTFCMTAIPNVMNFTSOANKTEVSGSEE 373  
202 YELRIVYLOOSDIEGGLGAPITVIVLCLFIAMLVFGVARGCKSGRAAYFLAPYV 261  
374 YKRYVFLKISAGIEEP-GEIRMPALCLFLAVAVIYASLAKIKTSRGVYFETAFFYV 432  
262 MTLFVTLTILPGADGILFEVTEPMARKLEGVYKSAVVOVFSLVCTGPIIMFSYN 321  
433 LVILLIRGVTLPGAGAGIWFVTEPKWEKLTNATWKADATQFEFSLAAMGGLTLLSSN 492  
322 GERNHIVYDAMIYVTLDTFTSFLSCCTIFGILGNLAVELNSEVDVYAGGTSIAFISY 381  
493 KFNHNCYRDVLIVCTNSATSHFAFVIFSVIGFMANRKNIEVADQ-CPGLAFVYVP 551  
382 DALAKTFQPOLFSVFEFLMSVLOGSSVALLSTFNITLAMAAPPVPTVYSAMT---CS 438  
552 EALTRPLSPFWAITFEFLMLTLGLDIMEFATIEIVISIDEPKRYLTRHPVFTLGGCI 611  
439 CGFLGLVYCTPGGYILELVHVGTFVLFCALISELAGVFWYGLCNDIEPMUGK 498  
612 CFEIFGPMITOGGIVMQLVDIYASALVITIAIFELVGIYVGLQRCEDIEEMMIGF 671  
499 KTGAVWRCKGVITTPAIVTVEFYLLASNNLVFGDNVYPTAGVSGYLMFLGTFVP 558  
672 QPNIFWYVMAFVPTLITFILCFEQWEMPTYG-SYRYPNMGVLMGLMGLACSVIMIP 730  
559 IGIGFSLYKRYRTGFSETIKKAFHKKPSMGSRPRERE 597  
731 IMFVYKMH-LAPGRFIERKLKLSQDPDWGFFLAQHNGE 768

RESULT 2  
US-09-795-693-24

Sequence 24, Application US/09795693  
Patent No. US20020068710A1  
GENERAL INFORMATION:  
APPLICANT: Glucksmann, Maria A.  
TITLE OF INVENTION: 20685, 579, 7114, 23821, 33894, and  
TITLE OF INVENTION: 32613, No. US20020068710A1 Human Transporters  
FILE REFERENCE: 35800/209292  
CURRENT APPLICATION NUMBER: US/09/715,693  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/185,906  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 624  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Pfam consensus sequence  
US-09-795-693-24

Query Match 30.6%; Score 1025.5; DB 10; Length 624;  
Best Local Similarity 37.6%; Pred. No. 1.7e-84;  
Matches 234; Conservative 94; Mismatches 201; Indels 93; Gaps 15;

QY 51 RMVYNNIEFLMSCIATSVGLNVMRFFPIAYONGGAFVLPVYVLLVGVKPVYLCV 110  
DB 1 RETWSGKLDVLSVGFVAVGLNVMRFPYLYCKNGGALILPLILFLIYAGIDPLFLELA 60  
QY 111 LGGFSSRNSVYK-----SIPAMGTGYAAGCGYLLSYVVICGLCYLLAMS 161  
DB 61 LGGYTRGSSITVYRKRIIDKKGICELFGIGVYASIVAFYIGIYVNIIVIAWALYLLFS 120  
QY 162 FQATLPALCOPEME--NCV-----PSDPTLAA-SVNNIIN---GSSAQILFLRY 207  
DB 121 FTTELPMATCNNSMNPNCVERERENSTNGSLAALSNNLTDYLERKSPVEEERWERY 180  
QY 208 LQ--QSDGIEGGLGAPITVIVLCLFIAMLVFGVARGVKS-SGRAAYFLAPYVIMT 264

181 LKLESSGIE-DLGEELRHELTCLLLAMIVYFCLMGVSGSKVYFPATFPYVLLIY 239  
265 LEITVILPGATDGLFEFVPPQMAKLELGWYSANVOVFSLVCTGPIIMSSYNGFR 324  
240 LILINGVILPGADGILFEVTEPMARKLEGVYKSAVVOVFSLVCTGPIIMSSYNGFR 299  
325 HNIRVAMIVYVTLDTFTSFLSCCTIFGILGNLA----- 357  
300 NNCYRDALIVYSINITSIFLAGFYFSLGPMANIVQEGVPENEKILLSVSLRDLPH 359  
358 -----YELNSEVD---VVC-----AGTSIAFTSYDAIAKTFQPO 391  
360 VNLSALRADYSVDIVISEVAESEFVLGLACLEDLKVQAGPGIAFAFYBEAATMLPLSP 419  
392 LFSVLPFLMSVLOGSSVALLSTFNITLAMDAPP-----RVPYVYMSAMTSCGFLGIY 447  
420 FMAVLFELMLTLTGLDQFGVGEITIALVDEPILIRKVRRELFILVCVISFLLGLEW 479  
448 CTGCGGYILELVHVGST-FLVFCALISELAGVFWYGLCNDIEFMUGKKGAYWRL 506  
480 VTEGGIVFVFLFYAASGFSLLFVVEFCIAVAMVYGDIDRFYDDITEMLGFRPGLYML 539  
507 CMGVTTPAIVTTFEVAL-----LASNNLV-FGDNVYPTAGVSGYLMFLGTFVP 558  
540 CMKFPVSLILFLFISIVQYGLKPLTYNNMIREADYVPMNANALGMLLSSMLCVP 599  
559 IGIGFSLYKRYRTGFSETIKKAFHKKPSMGSRPRERE 597  
600 LVIIYKLLSTEGDSLIERLOKA 621

RESULT 3  
US-10-156-239-24

Sequence 24, Application US/10156239  
Publication No. US20030036074A1  
GENERAL INFORMATION:  
APPLICANT: Glucksmann, Maria A.  
TITLE OF INVENTION: NO. US20030036074A1 Nucleic Acid Sequences Encoding Human Tr  
TITLE OF INVENTION: Atpase Molecule, A Human Ubiquitin Hydrolyase-Like Molecule, A  
FILE REFERENCE: 35800/247645  
CURRENT APPLICATION NUMBER: US/10/156,239  
CURRENT FILING DATE: 2002-05-24  
PRIOR APPLICATION NUMBER: 09/795,693  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: 60/185,906  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: 09/809,557  
PRIOR FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: 60/192,018  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: 01/808,568  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/191,790  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: 09/808,767  
PRIOR FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: 60/191,781  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 624  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Pfam consensus sequence  
US-10-156-239-24

Query Match 30.6%; Score 1025.5; DB 15; Length 624;  
Best Local Similarity 37.6%; Pred. No. 1.7e-84;





```

Db      65 PYCYONGGAGFLIPYCVMLFPGLEFFLEALGQYHRCGCLTLMKRICKALKGVYAI 124
Oy      137 AAGGCTILSYVYVIGCLLYL---AMSFQATLPMAIQOPMEN--CVPSPPTLAASVNN 191
Db      125 CMIDIMKMYNTIIGMAVYTLIASINSVLPMTSCDNEMNPLCTP-----VTSPT 179
Oy      192 ITNGTSSAQLYELFTVLOO--SDGIEGLGAPIMVLYCLFLAMLMVGVYARGVSSGK 249
Db      180 NPNSTPAKEFFERNVLEQHKNSGLD--DMGPIKPSLACVGVFLVYFSLMKGVRSAGK 238
Oy      250 AAYFLALFPYVMTLFTTITLPGATDGIIEFYTPQAKLLELGVWSAVTQVFFSLTV 309
Db      239 VVMTALAPYVLLILLAGVTLPGATEGIRYVLPPEHKKLONSQVMDAASQIIEFSLGP 298
Oy      310 CTGPIFMSYNGFPHNITRDAMIYTLDTFTSFLSGCTIFGILGNLAEYELNSEGVYVG 369
Db      299 GFGHLLAISYTKRNNKCYRDLITSSINCLTSFLAGVITFSLGYMAHVOKST--EEVG 357
Oy      370 AGGTSIAFISYPAIAKTFQDPSVLFELFMAVSGISSVALLSTFTNLAMDAPRY-- 427
Db      358 LEGGLVETIYVPEAIATMTGSVFMAIIEFLMLITGLDSTFGLEAVTALCDEVPRVIG 417
Oy      428 --PVTYMSAMTSCGFLGLVYC---TPGGQYILELVHYGTELVFCAISELAGVF 480
Db      418 RHREFAVL-----LFIYICAIPTTYGGVYLVLDLANYGPDLALFVFAEAGVC 471
Oy      481 MIYGLENCIDIEFLMGKKTGAYMRLCWGITPAIMTVFFYALLASNNLVFGDYVYPT 540
Db      472 WVIYGVDRSEVRYTMLGHTPGWEMTQWYSIPVFLVLEFVSVAHEML--GGEYTPS 530
Oy      541 AGVYSGYLMFLMGTFVPIGIGFSL 565
Db      531 WSIYGVWMTGTIVSCIPLYITIKL 555

```

## RESULT 8

```

US-09-919-039-378
; Sequence 378, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919, 039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 378
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID NO. US20030108871A1 5834958CD1
US-09-919-039-378

```

```

Query Match      27.8%; Score 932.5; DB 12; Length 614;
Best Local Similarity 33.8%; Pred. No. 4.6e-76;
Matches 208; Conservative 116; Mismatches 241; Indels 51; Gaps 14;

```

```

Oy      17 MEPRSSQISLPPA-----NNKKAALNDIDDTDLAEPPERAWWSNNIEFLNSCIATSYGL 71
Db      1 MDGKVAAYGEYPRAYSWVPEGEKLDQEDQVQK-----DRGWTAKMEFVLSVAGEITGL 56
Oy      72 GNVWRPFPIAYONGGAGFLVPYIVLLVGRPVYVLECVLQFSSRNSVYKW--SISPAWK 130
Db      57 GNVWRPFPIAYONGGAGFLVPYIVLLVGRPVYVLECVLQFSSRNSVYKW--SISPAWK 116
Oy      131 GTGYAAGAGCYILSYVYVIGCLCYLYLAMSFOALTPALIQOPEN--ENCVP--SPTPLAA 187

```

```

Db      117 GIGLASVIESYLANVYIIIIILAMALFYLSFSTSELPTWTCNNFWNTEHCTDFLNHSAG 176
Oy      188 SVNNITNGTSSAQLYELFTVLOOQSDGIEGAGAPIMVLYCLFLAMLMVGVYARGVSS 247
Db      177 TVTPENTSPYMEWERVRLGITGHI--DUGSLRWEALCILLAMVLCYFCIMGVKST 235
Oy      248 GKAAFLALFPYVMTLFTTITLPGATDGIIEFYTPQAKLLELGVWSAVTQVFFSL 307
Db      236 GKVVYFTATFPYLMVILLINGVTLPGAVQGIITYLAKDELRLKQPYQWMAAGTQIFFSF 295
Oy      308 TVTCGPIFMSYNGFPHNITRDAMIYTLDTFTSFLSGCTIFGILGNLAEYELNSEGV 367
Db      296 AICQGLATLALSGSYNNHNNCKYCDALCFLNATSFEVAGFVFSITGFMQEOGVPIE- 354
Oy      368 VGAGTSLAFISYPAIAKTFQDPSVLFELFMAVSGISSVALLSTFTNLAMDAPRY- 426
Db      355 VAESGPGIAFTAFPAVYVMPISQMSCLFELMLFLDLDSOPVCBECLVYASIDMPEQ 414
Oy      427 -----PVTYMSAMTSCGFLGLVYCTPGQYILELVHYGCT--FLVIFCAISE 475
Db      415 LRKSGRRRLTLTI---AVMC---YLIGLFLVTEGCMYIFQFDYASSGICLFLSLFE 468
Oy      476 LAGVYIYGLENCIDIEFLMGKKTGAYMRLCWGITPAIMTVFFYALLASNNLVFGDN 535
Db      469 VVCISWYVYVYADRFYDIEDMIGRWPVLKISWLEITGLCLATFLFSLKTYTPDKYNNV 528
Oy      536 YVYPTAGVYSGYLMFLMGTFVPIGIGFSLKRYRGTSEYIKKAF---HSKP----- 585
Db      529 YVYPTAGVYSGYLMFLMGTFVPIGIGFSLKRYRGTSEYIKKAF---HSKP----- 587
Oy      586 -----SMGPRSPRE 594
Db      588 LDGSAGRNFGPSPRE 603

```

## RESULT 9

```

US-09-815-923-14
; Sequence 14, Application US/09815923
; Publication No. US20020197644A1
; GENERAL INFORMATION:
; APPLICANT: Gill, Sarjeet S.
; APPLICANT: Ross, Linda S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A
; FILE REFERENCE: 023070-093800US
; CURRENT APPLICATION NUMBER: US/09/815, 923
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Manduca sexta
; FEATURE:
; OTHER INFORMATION: GABA transporter
US-09-815-923-14

```

```

Query Match      27.7%; Score 929.5; DB 11; Length 597;
Best Local Similarity 34.0%; Pred. No. 8.4e-76;
Matches 202; Conservative 121; Mismatches 228; Indels 43; Gaps 14;

```

```

Oy      17 MEPRSSQISLPPA-----QISLPPANNKKAALNDIDDTDLAEPPERAWWSNNIEFLNSCIATSYGL 71
Db      1 METKNDSDSDIELSAOSGKNKPS-----DAVAKSNLEBRGSMASKLDFIISVIGLAIGL 55
Oy      72 GNVWRPFPIAYONGGAGFLVPYIVLLVGRPVYVLECVLQFSSRNSVYKW--SISPAWK 131
Db      56 GNVWRPFPIAYONGGAGFLVPYIVLLVGRPVYVLECVLQFSSRNSVYKW--SISPAWK 115
Oy      132 TGYAAGAGCYILSYVYVIGCLCYLYLAMSFOALTPALIQOPEN--ENCVP--SPTPLAA 185
Db      116 TGYAAGAGCYILSYVYVIGCLCYLYLAMSFOALTPALIQOPEN--ENCVP--SPTPLAA 175

```

QY 186 AASVNNIT-----NG--TSSAQJ-----YFLRTVLOQSDGIEGGLGAPIMVYLCEPIA 232  
176 MSSLDGDMSTFCTLNGRNVSKAVLSIPKKEFERERALDISSGIE-HIGNIMELAGTLLV 234  
QY 233 WMAVEGVAVGAVSSGGAALFLLIPYVMITLFTITIIIPGATDGLFVTPQWAKLE 292  
235 WVLCEFCIMGVKVRMTGKVVETLLIPFELLTVLLIRITLPGAMEGKIFVMPMMSKLE 294  
QY 293 LGVYSAVTQVFEFLVCTGPIIMISSNGFRHNITVDAMIVTLDTFTSELSGCTTLEGI 352  
295 SEWIDAVTQIFSFYSGIGLGLTILVANSYNKFTNNVYDALIVCSNSTSMFACQVTFVSV 354  
QY 353 LGNLAELNSEVDVVGAGGTSIAHISYDPALAKTFOPLFVFLFMSVLAGSSVAL 412  
355 VGEFAHQORPVAE-VAASPGJALILAYPSAVILQPLGAPLMSCLFFEMLLIGDSQFCT 413  
QY 413 LSTFNTLAMDAPRV-----PTVTKAMTSCGFLGLGYCTPGQYILEVDHIG-GTFL 467  
414 MEGITAVIDEMPLLRREITLITATCITISYLVGSCISEGMYEQILDSYAVSGFC 472  
QY 468 VLEFCAISELAGVWYIGLENLCLDIEMLKKTGAYRCLMGVITPAIMTTFEYALLAS 527  
473 LFLIFEFCYSISAFVNNFYCIKEMIGYPIIMKFCWGTTPAICISVFLFNLYOM 532  
QY 528 NMLVEGNYVYPAGYVSYLMLFMGMTFVPIGIGFSLYKRT--GTSETIKK 579  
533 TPICY-MNVEYPMWSHAQWETALLSMCLIP---GYMILMYRVTPTGMOEKFRH 582

## RESULT 10

US-09-815-923-10  
; Sequence 10, Application US/09815923  
; Publication No. US20020197644A1  
; GENERAL INFORMATION:  
; APPLICANT: Gill, Sarjeet S.  
; APPLICANT: Ross, Linda S.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A1  
; FILE REFERENCE: 023070-093800US  
; CURRENT APPLICATION NUMBER: 03/09/815, 923  
; CURRENT FILING DATE: 2001-03-23  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 727  
; TYPE: PRT  
; ORGANISM: Manduca sexta  
; FEATURE:  
; OTHER INFORMATION: neurotransporter transporter encoded by inebriated  
US-09-815-923-10

Query Match 26.3%; Score 882; DB 11; Length 727;  
Best Local Similarity 33.9%; Pred. No. 2.3e-71;

Matches 188; Conservative 108; Mismatches 219; Indels 40; Gaps 14;

QY 28 PPAANKALDINDTDLEAE-----PERMWSNIEFLMSCIATSVGLGVNMFPE 79  
138 PRPEDAS-----HESDEEEELAAVHLKRPRFWANKIOFVLACGYISGLGVNMFPE 193  
QY 80 IAVONGGAPLVYVYLLVHPYVYLECVLOFSSRSNVKWS-ISPAMKGTGYAQA 138  
194 LCYKSGGAPLIFYFILLICGPNLEMLAIGQYTHAGPIGALSQICPLFKAGLASV 253  
QY 139 GCGYIIISYVIGLCIYLAIFGATLPAICQPER--ENC--VPSPITLAASVNTT-- 193  
254 ISFLMSTYAVIILAMALYFTFTFTEVPWASCSNRNIDQCWPN-----HNTKP 305  
QY 194 NGSIS-AQLFLRTVLOQSDGIE--GGLGAPIMVYLCEPIAMLVGAVGVKSSGKA 250  
306 NGSOTPEORFERKVLNMSAGIHYIGMR---WELACVLCVAVLYIFALMKSIKSAKY 362

QY 251 AYFLALPYYVMTLFTITIIIPGATDGLFVTPQWAKLELGWYSATQVFEFLVCT 310  
363 RYITTLPLFLLIYFELGRSLTLDGADGLAFKPRPDWELLKQSRPWMAASQIFNSGIA 422  
QY 311 TGPLMSSYNGFRHNITVDAMIVTLDTFTSELSGCTTIGLGNLAELNSEVDVYGA 370  
423 FGSIMFASYNRPNLHPTVAVTLVNAITSLVIGFIFPATIGNIAFEONTFVKDI-A 481  
QY 371 GGTSLATSPDAIAKTFQOLFSVLEFLMMSVLAGSSVALSTFTMLMDAPF--RV 427  
482 DSPGLFVVPQALAKKPASQMAVLEFEMFLCGLNSQPAIVEVVTSTQDGPDMIRK 541  
QY 428 PTVY---MSAMTSCGFLGLGYCTPGQYILEVDHIGGTFTVLEFCAISELAGVWYIG 484  
542 RLVYHELVLVLCVAVSLDGLPHIHSGIYFQMDVYAAASLTITLAFEEVVALMIFYG 601  
QY 485 LENICLDIEMLKKTGAYRCLMGVITPAIMTTFEYALLASNMLVEGNYVYPAGYV 544  
602 VGRLSRNKQMTGQPSLYFRFCWMLASPALLLALWASVVDYTPPSY-RQYQYPAWAQA 660  
QY 545 SGYLMFLGMTFVPI 559  
661 LGWIMASLSLICIV 675

## RESULT 11

US-09-815-923-6  
; Sequence 6, Application US/09815923  
; Publication No. US20020197644A1  
; GENERAL INFORMATION:  
; APPLICANT: Gill, Sarjeet S.  
; APPLICANT: Ross, Linda S.  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A  
; FILE REFERENCE: 023070-093800US  
; CURRENT APPLICATION NUMBER: US/09/815, 923  
; CURRENT FILING DATE: 2001-03-23  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Manduca sexta  
; FEATURE:  
; OTHER INFORMATION: proline transporter  
US-09-815-923-6

Query Match 25.9%; Score 869; DB 11; Length 556;  
Best Local Similarity 33.6%; Pred. No. 2.4e-70;

Matches 188; Conservative 103; Mismatches 235; Indels 34; Gaps 11;

QY 54 WSNNIEFLMSCIATSVGLGVNMFPEPIAYONGGAPLVYVYLLVAGRVYILECVLQ 113  
9 WGSQLEVLISLCLDIAVAGIGMLMRPVLQYRNGGAPLIFYELLICGIPLVYLETLGQ 68  
QY 114 FSSRSNVKWSISPAMKGTGYAQAAGCYILSYVYVIGLCIYLAIFGATLPAICQ 173  
69 FASAGCISVNNINPLFRGAGYAVIYLVNVAISIFSAIMSPILIXIYHSMKSPPLPMSG 128  
QY 174 EME--NC--VPSPITLAASVNTTNGTSSAQLYFLRTVLOQSDGIE--GGLGAPIMVYL 227  
129 SMTVIVCTEITGNSFSTSTTPEDE---YFHRLLQVSNINIHIGSIYAVFVW--- 182  
QY 228 CLFIAMLVNGVAVARVYKSSGKAAYFLALFPYVMTLFTITIIIPGATDGLFVTPQW 287  
183 CNLICMLVYLICNGKSVKIVFTVLPYVVLVLEVRGILTPGAMGIMFYILPDW 242  
QY 288 AKLLELGWYSATQVFEFLVCTGPIIMSSYNGFRHNITVDAMIVTLDTFTSELSGC 347  
243 AQLAKRVADAATQIFSLGPGMGGLVMSNFKHYNNLRSSMILPIYNSATSTIMAGF 302



[illegible]

	Query Match	25.4%	Score 852.5;	DB 10;	Length 671;
	Best Local Similarity	32.6%	Pred. No. 9,7e-69;		
	Matches 193;	Conservative 114;	Mismatches 256;	Indels 29;	Gaps 12.
OY	22	SSQSLPNNKKAALDINDDD-----LEAPEPMWMSNNIEFLMSCIATSVGLGNV	74		
	1:	1:	1:	1:	1:
Db	64	STSHSIDRNEPLAALGGILTPKEGRVALRRSSMWDRKATMEFLAVVGAVDLGNL	123		
	1:	1:	1:	1:	1:
OY	75	WPEPPIAYONGGAGFLVPPYVLLVJCKPPYILECVYEGOSSRNSKYW--SISPMKGTG	133		
	1:	1:	1:	1:	1:
Db	124	WPEPVCYKHHGGAFILPYFLMILITGGLPMFYELVAGHNSGCYSIRKRCPLFRIG	183		
	1:	1:	1:	1:	1:
OY	134	YQAAGCCTIISTYVVICGLCYLYLAMS---FOATLPMAICQPEWENCVPSDPTLAAY	189		
	1:	1:	1:	1:	1:
Db	184	YGCICCFIPIFYFNATIAQAVPAIVASYISKINDSEVPMAISCGNPNNTPRCDDLNTVTS	243		
	1:	1:	1:	1:	1:
OY	190	NNITNGTSSADLYFLRYL--QOSDGTGGGLGAPIVLYCLFIAMLVGVAARGVKS	247		
	1:	1:	1:	1:	1:
Db	244	RNGPIPLTPSEYYLYLVLEVOKSTGPD-DLGGVKTSMACVALLAVIMVYFALMGKQSS	302		
	1:	1:	1:	1:	1:
OY	248	GKAALFLLPYYVMILFETITIIIPATDGLILEFTPMQAKLLBELGWYASVYQFESL	307		
	1:	1:	1:	1:	1:
Db	303	KTIWVMTATAYIILISLLIRGLLPLPAKNGKLYYVTPFKELKAPWASAAADNIFPSL	362		
	1:	1:	1:	1:	1:
OY	308	TYVCGPIIMSSYNGEFRHNTYRDAMVYTLTDTFTSFSCITFGILGNLAEVLSQVDY	367		
	1:	1:	1:	1:	1:
Db	363	GGEGEVLALSSYDQFENNCCYRDAAVYLSIINCATSPFSCVVFSTLGTMSLLTKNPINEV	422		
	1:	1:	1:	1:	1:
OY	368	VGAGGTSLAFSTYDAIAKTFQPLSVLEFLIMSVLIGGSSVALLSTFNTLAND--AFP	425		
	1:	1:	1:	1:	1:

```

Db      423 VGEHDASLILITVYPQALAMDSCSFSSFFFMILITLGDSFGIAEAFITGCDESRFL 482
QY      426 RVPTVYVMSAMTCSGCGELLGLVYCTGBOGYILELVHYGTFLVFCAISELGAVFYWIGL 485
Db      483 SKNKWEVIALICITYFLSFPAISGOGYVIFDLEGVSLVLFYTCEMANCWMYYV 542
QY      486 ENLCDIEFMLKKTAYRLCWGYITPAIMTVFFYLLASN-NLVGDNVYPIAGYV 544
Db      543 DQSKDIRAMLGYPGYIAVWCM-RCSPFISVIVIMTVNSSKPIOMASYTFPMWSVI 601
QY      545 SGYLMLFGTVPIIGSLKYFGTFSEFIKAFSRDSKMOPRSRPR 596
Db      602 LGMFRLLSYLAIPV---FAILYLSGT-GLYRFR----WA-ITPQQR 643

```

```

RESULT 13
US-09-843-598-7
; Sequence 7, Application US/09843598
; Patent No. US20020010944A1
; GENERAL INFORMATION:
; APPLICANT: Horvitz, H. Robert
; APPLICANT: Ranganathan, Rajesh
; TITLE OF INVENTION: DESERT GENES, PROTEINS, AND MODULATORY
; TITLE OF INVENTION: COMPOUNDS
; FILE REFERENCE: 01997/525002
; CURRENT APPLICATION NUMBER: US/09/843,598
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/200,549
; PRIOR FILING DATE: 2000-04-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 671
; TYPE: PR1
; ORGANISM: Caenorhabditis elegans
; US-09-843-598-7

```

Query Match	25.2%	Score 844.5	DB 10	Length 671
Best Local Similarity	32.4%	Pred. No. 5.2e-66		
Matches 152	Conservative 114	Mismatches 257	Indels 29	Gaps 12
QY	22	SSQISLPANNAKALDINIDDD-----LEAPEPPMWSNNIEFLMCSIMTSGVIGNV	74	
Db	1	1	1	1
QY	64	STSHSIDNEPIALGLGLPTKEGRVAALRRSSMRDKATMTAEFLILAVGYRADLGN	123	
Db	1	1	1	1
QY	75	MEPPPIAONGGAFLVYVVLVLLVKRPVYLYECVLGQFSNNISYKVN--SISPMKGTG	133	
Db	1	1	1	1
QY	124	WEPPVSVCKRHGGGAFLDPYFLMILIGLPMFYMELVIGQFHRSQCSYIRKRCPLFRIG	183	
Db	1	1	1	1
QY	134	YAAAGCCYIILSYVVICGLCYLYLAMS----FOATLPMACIOPHEWNCVPDPFLASV	189	
Db	1	1	1	1
QY	184	YGICICTEFIAPFNAILAAQAVYFAIVLSKIMDSEVPANASGCPNPNPRCSDLIANTIS	243	
Db	1	1	1	1
QY	190	NNITNGTSSADLYFLRYVL--QOSDGTGGGLAPIVLYLCLFIAMLVFGVARGVSS	247	
Db	1	1	1	1
QY	244	RNGIPLTTPSEYLYLVYLEVKSTGTFD--DLGCVKTSMAVCLLAVEMIVYFLMKGPOSS	302	
Db	1	1	1	1
QY	248	GKAAYFLALFYVVMIFLFTYITLLPSPATDGIILFETTPQMAKILEGVYSAVTVQFSL	307	
Db	1	1	1	1
QY	303	GKIYVVTATAYIILITILLRGLLPEAKNGLYVYTPDEKLIKDPAYMSAAADIFPSL	362	
Db	1	1	1	1
QY	308	TWCGEILIMSSNGFPHNIIYRDAMIYITLIDFTFSLSCCTIFGLIGLNAVELNSEVDV	367	
Db	1	1	1	1
QY	363	CGPGCVILALSYNDPENNNCYRDPAVYIISINCATSEFCVVFSTLGYMSLITKNPINEV	422	
Db	1	1	1	1
QY	368	VGAGGTSLATISYPDATATFOPLQSVLFEFLMVAIVIGSSVALISFTNTLAND--AFP	425	
Db	1	1	1	1
QY	423	VGEHDASLIEIFYPOALATMDYSCFMSFIEFVWLITLIGIDSPFAIEAFITGFCDESEFL	482	
Db	1	1	1	1
QY	426	RVPPIVYSAMMCSGCELLGLVYCTPGCVITLLEVDHGTGTYLVEFCALISELAGVFWIYL	485	
Db	1	1	1	1
QY	483	SKNRKRVLYICIIYFLSPAISVSGGVPIPLDEYGSVLSLFTVCCIMAIACWCFEY	542	
Db	1	1	1	1

QY 486 ENICLIDIERMKGKTKGATYRLCQVITPAIMTVEFYALLASN-NLYPDNYYVPTAGVY 544  
Db 543 DQFSKIDIRAMLGFPYIGIYVWCM-TQSSVFISVIFIMTYNSSFEDIQASQTFPMWSVI 601  
QY 545 SGYLMFLGATFVPIGIFSLYKRIQTESETIKKAFHSKPSWGPSPRRER 596  
Db 602 LGMFLRLSLVALIPV---FAITILLGT--GTIXEFR---WA-ITPQOR 643

RESULT 14  
US-09-815-923-12

Sequence 12, Application US/09815923;  
Publication No. US20020197644A1  
GENERAL INFORMATION:  
APPLICANT: Gill, Sarjeet S.  
APPLICANT: Ross, Linda S.  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Use of insect cell Membrane Transporters as No. US20020197644A1  
TITLE OF INVENTION: Target Sites for Insecticides  
FILE REFERENCE: 023070-093800US  
CURRENT APPLICATION NUMBER: US/09/81, 923  
CURRENT FILING DATE: 2001-03-23  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 12  
LENGTH: 676  
TYPE: PRT  
ORGANISM: Manduca sexta  
FEATURE:  
OTHER INFORMATION: orphan transporter  
US-09-815-923-12

Query Match 24.8%; Score 833; DB 11; Length 676;  
Best Local Similarity 33.3%; Pred. No. 5.8e-67;

Matches 197; Conservative 95; Mismatches 204; Indels 96; Gaps 14;

QY 44 LEAEPPE--RWVSNIEFLMCIATSVGLGNVRRFPFIAYONGGAFVLPYVILLVG 101  
Db 43 LQRAPEEDRAMSGKLOFPLSTIGTSYGLGNIMRFPPLCQNGGAFLLPFLIMLVLEG 102  
QY 102 KPVYVLECVLGQFSSRSNVKWA-SISPMKGTGYAQAAGCYILSYVVICGLLYLAM 160  
Db 103 IPLVYIEAMIGCKMRLGSLGVNTHPWLGGIGISSCVLTFLVLYNVITWVFFLEN 162  
QY 161 SFOAT--LPWATCOPEWENCVPSPDLAASVNNITNGT-----SSAQLYF-LRTVLQ 209  
Db 163 SIRLTADQLPNAHC-----HYD-----NCTAABECNKSATVYFVYREALD 203  
QY 210 OSDGIEGIGAPIMYLVLCFLTAMLVGVVARGVSSGKAAYFLAFLPYVMTLFTT 269  
Db 204 ASPSIDP-GVPRMVIYLLLAWITVFIYMKGIQSSGKVYFSLFPYAVLTIFFVYG 262  
QY 270 IILPATGILFFVYPOKAKLLEQVWYSAYTOVEFSLVCTGPTIMSSYNGFRHNHYR 329  
Db 263 IITLPOSSGILHMYPKLEKLDLPVWMDAATOVYFSGILAFSGIAPNPNNCYR 322  
QY 330 DAMIYTTLDTFSELSGCTIEGILGNLAY-----ELNSEV--- 364  
Db 323 DVLLVSYCNALTAIYASVIESILDFKAYTVENCIVKEIKVLAHLHIGFTLNSTADY 382  
QY 365 -----GDVV-----GAGTSIAFISYDPAIAKTQPOLFSVLEFLM 400  
Db 383 REQPRFLNLTALNALNLGCTMRGLLEBAECTGATFVFTQALIKLPAPFWSTIIFLM 442  
QY 401 MSVLIGSSVALLSTF--NTLANDPFRVPTVYMSAMTSCGFLGLVYCTPGQYIEL 458  
Db 443 LLSLIGSQIGIMEGLCTIFDIDFKRLSKPVITGVCTCFEFGILFTTGAGGYWLMK 502  
QY 459 VHYGGTFVLFCALSELAVFPIYGLEMLCIDIEFMLGKKTGAYWRCLQWVITPAIMT 518  
Db 503 PFSFAGTIGLVYVALLKMAVITIGHEKFTNDIYEMGYRPGIYQWVYWRVGAIVTC 562  
QY 519 VFFYAL--LASNNLVFGD-----NIVYPTAGVSGYLMFLGIMFVPI 559

Db 563 ILLSLVEMLINPMPYGAWMNDEGRVITPYPTVLYVIAVLMILAGVLPV 614

## RESULT 15

US-09-741-149-2

Sequence 2, Application US/09741149  
Patent No. US20020031800A1  
GENERAL INFORMATION:  
APPLICANT: Li, Zhenya et al  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
AND USES THEREOF  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: CLO00780  
CURRENT APPLICATION NUMBER: US/09/741,149  
CURRENT FILING DATE: 2000-12-21  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 730  
TYPE: PRT  
ORGANISM: Human  
US-09-741-149-2

Query Match 23.6%; Score 790; DB 10; Length 730;

Best Local Similarity 29.3%; Pred. No. 5.2e-63;

Matches 190; Conservative 111; Mismatches 218; Indels 130; Gaps 11;

QY 2 NCGVNGGSESPKPEKRRSQISLPANNNKAALNDIDDTLE--AEPERWVSNNI 58  
Db 26 NEDAADARTISLIVDQGEK-----DTVEEGSEVEDERPRAMNSKL 68  
QY 59 EFLMSCIATSVGLGNVRRFPFIAYONGGAFVLPYVILLVGKFPYVLECVLGQFSSRN 118  
Db 69 QYLAQVGSVGLGNVRRFPYLCQKNGGAGAYLLPILIMVIGIPFLFELSLVGORIRG 128  
QY 119 SVAVWS-ISPAMKGTGYAQAAGCYILSYVVICGLLYLAMFOATLPAIC----- 171  
Db 129 SIGVWVYISPKLGIGISFACVCVCFYALVLYNIGSLSEFSOSFQDPLPMQCPYKNA 188  
QY 172 -----OPEWENCVPSPDLAASVNNITNGTSSAQLYFRTVLVQOSDGI--EGGLAPVY 224  
Db 189 SHFVEPECGQ-----SSATYVWYREALNMISSISEGLN--WK 227  
QY 225 LVLCFLTAMLVGVVARGVSSGKAAYFLAFLPYVMTLFTTITLPGADGILFFVT 284  
Db 228 MTICILAAWVWCLAMIKIGIOSGKITIYESSLEPPYVLLICFLRAFLLNGSIDGRHMT 287  
QY 285 POKAKLLEQVWYSAYTOVEFSLVCTGPTIMSSYNGFRHNHYRDMIVTTLDFTSFL 344  
Db 288 PKLEIMLEPKVWREATOVFFALGLGFGVIAFSSYNKNDNCHPDAVLVSFINFTSVL 347  
QY 345 SGCTIFGILGNLAYELNSE-----GAGTSIAFISYDPAIAKTQPOLFSVLEFLM 401  
Db 348 ATLVVFAVLGFKANVINKECITONSETIMFKMKNISODIIPHNILSTVYAEADYHLY 407  
QY 366 DVV-----GAGTSIAFISYDPAIAKTQPOLFSVLEFLM 401  
Db 408 DIIOKVEEFPALHNSCKIEELNKAVOGTGLAFLAETHTHPASPFWSMVFLML 467  
QY 402 SVLIGSSVALLSTFNTLADAFPRVPTVYMSAMTSCGFLGLVYCTPGQYIELVDH 461  
Db 468 VNLGSGMEGTIEBIYPIYDIF-KVRKELLVYICCLARICIGLIFQSRGNYFMFDD 526  
QY 462 YGTFVLVFCALSELAVFPIYGLEMLCIDIEFMLGKKTGAYWRCLQWVITPAIMT 521  
Db 527 YSATLPLIVYLIENINAVCEVYIGDKFMEDLKMLGAPAPRYYYMKYITSLMLSL 586  
QY 522 YALLASNNLVFGDN-----YVPTAGVSGYLMFLGIMFVPI 559  
Db 587 ASVWNKGLSPGYNAMIEDKASEEFLSYPTWGLVVCVSLVFAILPVPV 635

Mon Jul 28 08:01:33 2003

Search completed: July 21, 2003, 09:37:48  
job time : 60 secs

us-09-991-458-2.rapb

Page 9

**THIS PAGE BLANK (USPTO)**